

[illegible]

```

RESULT 7
US-09-305-384-5
: Sequence 5, Application US/09305384
: Patent No. 6242218
:
: GENERAL INFORMATION:
: APPLICANT: Treco, Douglas A.
: APPLICANT: Heartlein, Michael W.
: APPLICANT: Selden, Richard F
:
: TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY

```

```

; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: US 60/084,649
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 6235
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-305-384-5

Query Match          6.4%; Score 169.6; DB 4; Length 6235;
Best Local Similarity 54.4%; Pred. No. 5.1e-25;
Matches 442; Conservative 0; Mismatches 344; Indels 26; Gaps 4;

QY 1869 aaaaattagctggcgctggtggcattgcatcaccaatccccagctactctggggaggctgaggc 1928
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 207 aaaaattagcaggcatggtggcgagcactgtatcccgctactcagggagctgaggc 266
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1929 atgagaatcgttgaaccgggaggcgagatgttcagtgagccgagagcgcgccactgca 1988
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 aggaagatcacttgaaccggaggcgagggttcagtgagctgagatccaccactgca 326
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1989 ctccagcctggactacagagcgagactctatctcaaaaaaaataaaaaaaagtaac 2048
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 ctccagcctgggtgacagacgaagactctatctcaaaaaaaataaaaaaaataaaaaatt 386
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2049 ttagggtcagggt-----gtcctctgtttatcactgagaccgtgcgccggttatgggt 2102
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 agccaggcattggtagtcgacacctctagctcagctactcagggagctgaggtgggagga 446
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2103 tgtaccagaaagcaagatttc-actatgcacatattcacgcgtcacccctagcattgaag 2161
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 tcacttgaacctggggcagtcgaagctcacgtgagcgaagatcatgcctacactccag 506
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2162 ccagcctgtagcctgaaagccttgccttggggcaggctcttccccaaaatcgagacac 2221
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 cctgggcaacagagagagaccctgtctcaaaaaataataataaagaaaaaaacacag 566
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2222 gaaggtgcaaaagtgaagctgccagctcttgcaaaagatgttaactgtcacgaagccacga 2281
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 567 ctctgtttatgtctcctcctggtccatcacactactatgtatatagtttgcaaacacaaaga 626
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2282 gtgacgggagagctgtccacatttcgggaagt---ggctatgtgaggaagcggggagagc 2338
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 627 tceagatagtcgaatttttaggctgtggcgctatggtctctgtcacaatcactctgcc 686
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2339 gggccccctagagataagagacaaatcataaggggagagata-----tcaga 2382
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 ctgtcttctagcacaaaagcagctataaacaaatacatacagtaatttttatagacatc 746
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2383 gaaaatcgtgaaggggagcagatggtgtcagagaaataggctgacactcgagagactggc 2442
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 747 gagatttgaatttcattgatttttacattttataaaataactcttttcaaaaattttccc 806
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2443 agaagctttcagaaaaaccactggcgctgggcagctggccttaggcctgtaatcccaagc 2502
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 ctacaccattaaaagtgtaaaagccggcagcgcgccatcgctcacgcctgttaattccagc 866
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2503 actctggagagctgacacagctgaaactcttgaggtcagagagtlccagaccagcctggcc 2562
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 867 actctgggaggtcaggtgggcagatcacttgcagatcaacagttcgagaccagcctggcc 926
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2563 aacatggtgaaacccccatctctacagaaaaataaaaaattagccagcgctggtgaccaa 2622
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 927 aacatagcaaaaccccatctctactaaaaataaaaaaattagctgggcatagtggtgcac 986
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2623 gacctagaatccagctactcttggggagctgagg 2654
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 987 acctgataccagctactcttggagagctgagg 1018

```

```
RESULT 8
US-09-305-384-1
; Sequence 1, Application US/09305384
; Patent No. 6242218
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Selden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/017001
; CURRENT APPLICATION NUMBER: US/09/305,384
; EARLIER FILING DATE: 1999-05-05
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-384-1

Query Match          6.4%; Score 169.6; DB 4; Length 6679;
Best Local Similarity 54.4%; Pred. No. 5.1e-25;
Matches 442; Conservative 0; Mismatches 344; Indels 26; Gaps 4;

QY 1869 aaaaattagctggcggtggcgtgctatcccaatccagctactggggagcgtgagcc 1928
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1929 atggaatcgtctgaacggggagagacatgttgacgtgagccgagacggcgcactgca 1988
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 aggaatacactgaaccagggagggcggtggtgagtgagtgagatcacaccactgca 345
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1989 ctccagctggactagagagagactctatctcaaaaaaaataaaaaaaagtaac 2048
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 ctccagctgggtgacagagcaagactctctcaaaaaaaataaaaaaaataaaaaatt 405
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2049 ttagtgaggggt-----gtcctgtgtatcactgagacgtgcccgtgtatgaggt 2102
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 agccagcgatggttagtcacaccccttagctcagctactcagggaggtgaggtggagga 465
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2103 tttaccagaaagcaagtattc-actatgcacactattcacgcctcacccctgagcattgaag 2161
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 tcaattgaacctggggcagtgcaaggctacagtggcgaagatcatgccaactacactccag 525
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2162 caagcctgtagcctgaaagcccttctgttgaggcaggtcttcccaaaaatgcagacac 2221
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 526 cctgggcaacagagagagacccctgtctcaaaaaaaataataataaagaagaaaaacag 585
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2222 gaaggtgcaaaagtgaagctgcagctccttgcaaaagatgaactgttcacgaagccacga 2281
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 ctctgtttatgtctcctggtccatcacatactactatgtatagtttgcaaaactcaaga 645
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2282 gtggcagggagagctgtcccaactttgcgaagt---ggctatgtgagcagggggagcc 2338
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 tccagatagtaatttttaggctgtg999cgtgtgctctgtcacaatcactctgcc 705
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2339 ggggtcccttagatagaagacaatcataaggaggagata-----tcaga 2382
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 706 ctgtctttctagcaaaaagcagctataaacaataacatacatgaattttttatagaacatc 765
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2383 gaaatcgtgaagggagagagatggtgtcaagagaataaggctgaccctcgaagagactggc 2442
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 766 gagatttgaaatttcataatgattttttacattttataaataaatcttttttaaaattttccc 825
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2443 agaagctttcagaaacacactggcgagcgtggtggcacagtggccttaggccttaattccagc 2502
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 826 ctaaccatttaaaagttaaaagccggccagcgccatcgccagcctgtaattccagc 885
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 2503 actttggagggtgacgcaggtgaatcacttgaggtcagaggttccagaccagcctggcc 2562
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 886 actttggagggtgaggtgggcagatcacttgagatcaacagttcagaccagcctggcc 945
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2563 aacatgggtgaaaccccatctctacagaaaataaaaaattagccagcggtggtggcaca 2622
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 946 aacatagcaaaaccccatctctactcaaaaaataaaaaattagctgggcatagtggtgcac 1005
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2623 gctagaatcccgactctacttgaggcgtgag 2654
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1006 acctgtgattcccgactctacttgaggcgtgag 1037
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-08-480-784-20
; Sequence 20, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-480-784-20

Query Match          6.2%; Score 165.4; DB 1; Length 6769;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

Qy 1869 aaaaatagctggcggtggcgatgccacacatccacgactgagggaggtgagc 1928
Db 160 AAAAAATTAGTGGTGGTGGCAGCTACCTGTAAATCCAGCTACTCGGAGGCGGAGGC 219
Qy 1929 atgagaatcgttgaacgcggggagcgagatgttgcagtgagccgagcgccactgca 1988
Db 220 ACAAGAATTGCTTGAACCTAGGACGCGGAGGTTGCAGCGAGCCCAAGATCGCGCCACTGCA 279
Qy 1989 ctccagctgagactacagagcagactctctcaaaaaaataaaaaaagaat-aa 2046
Db 280 CTCCAGCTGGGCGGTGCTTCCCTGTATCCACAGCTTTGGGAGGCCAAGCGCGGTGGA 339
Qy 2047 acttagtgaggtgtcctctgttattcactgagaccgtgcccgggttatgagtt--- 2103
Db 340 AGCTGGGCGCAGTGGCTTTCCTGTATCCACAGCTTTGGGAGGCCAAGCGCGGTGGA 399
Qy 2104 -----gtaccagaaagaagtattcactatgcacattatccacgtccacctgagcatt 2157
Db 400 TCTCCAGTCTTAGAGTTCAAGACCCAGCCTAGGCAATGTGTGAACCCCATCTACAAA 459
Qy 2158 gaagcagcgtgtagcctgaagccttctgtttagagcgaggttttccc-----c 2208
Db 460 AAATACAAAAAATTAGCCAGGCGATGGTGGCGTCCGATGTATCCAGCTCTTGGGAGGC 519
Qy 2209 aaatgcagacacgaagtgcaagtgcaagtgcaagtgcaagtgcaagtgcaagtg 2268
Db 520 TGAGGTGGGAGGATCACTTGAACCCAGGAGACAGAGTTGCAGTGAACCGAGATCACGCC 579
Qy 2269 acgaagggccagagtgccagggagagctgttccacatttgcggaagtggctat----- 2321
Db 580 ACCACGCTCCAGCCTGGGCAACAGAAAGACTCTGTCTAAAAAATAACAAATAAAATAA 639
Qy 2322 -----gtgagacggggagggcggtcccttagagataagagacaatacgaaggagat 2376
Db 640 AAGTAGTCTTCACAGTACCAGCATTCATTTTCAAAAGATATAGAGCTAAAAAGGAAGGA 699
Qy 2377 atcagaaaaatcgtaaaggggagcagatggtgtgcaagagataagctgacccatcgaagg 2436
Db 700 AAAAAAAGTAATCTGGGCTTTTAAATACCTCTGTTCTATATACTAAATGTTCTTAGGAGTG 759
Qy 2437 actggcgaagccttcca-----gaaacacactggagcggctgggacagtgagcagtg 2483
Db 760 CTGGGCTTTTATTTCTCATATTTATCTCTTTTAAAAATGTTATTGGCAGGACGCGTGGC 819
Qy 2484 ttaggctgtaataccagcacttggagcgtgagcaggtggaatcacttgaagtcagga 2543
Db 820 TCATGGCTGTANTCCAGCATCTGGGAGGCGGAGGAGGAGATCACCTGAGGTCAGGA 879
Qy 2544 gtccagacacgctggccacatggtgaaaccccatctctacagaaaaataataaatta 2603
Db 880 GTGTGAGACCAACGCTGGCCACATGGCGAAACCTG-TCTCTACTAAAAATACAAAAATTA 938
Qy 2604 gccagggcgtggtgacaaagcctagaatcccgactcttgggaggtcagg 2654
Db 939 ACTAGGCGTGTGTACGCGCTGTAGTCCCGAGCTACTCGCGGAGGCTCAGG 989

RESULT 10
US-08-483-553-20
```

```
; Sequence 20, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,553
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6769 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-483-553-20
```

Query Match 6.2%; Score 165.4; DB 1; Length 6769;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

Qy 1869 aaaaatagctggcggtggcgatgccacacatccacgactgagggaggtgagc 1928


```
Db 160 AAAAAATTAGCTGGGTGTGGCAGCTACCTGTAAATCCAGCTACTCGGGAGCGGAGGC 219
QY 1929 atgagatcgtctgaacgggagagcagatgttcagtagccgagacgcccactgca 1988
Db 220 ACAAGAAATTCCTTGAACCTTAGCAGCGGGAGGTTCAGAGCGCAAGATCGCCACTGCA 279
QY 1989 ctccagcctggactacagagcagactctatctcaaaaaaaataaaaaaagt--a 2046
Db 280 CTCCAGCCTGGCCGTAGAGTGAAGTCTGTCTCAAAAAGAAAAAGTAATGTCTTCT 339
QY 2047 acttagtgaggggtgcctctgttattcaactgagacogtgcctccggttatgaggt--- 2103
Db 340 AGCTGGCGCAGTGGCTCTTGGCTGTAAATCCAGCACTTTGGGAGGCCAAGCGGTGGA 399
QY 2104 -----gtaccagaagaagatttcaactatgcacactattcacgcgtcacctagcatt 2157
Db 400 TCTCGAGTCTTGAAGTTCAGACCGCTAGGCAATGTGGTGAACCCCATCGCTACAAA 459
QY 2158 gaagccagcctgtagcctgaagaagccttctgttggagggcaggtcttctccc-----c 2208
Db 460 AAATACAAAAATAGCCAGGCATGTGGCGTGGCATGTAGTCCAGCTCTTGGGAGGC 519
QY 2209 aaaaatgcagacacgaaggtgcgaagtgaagctgcagcttctgcaaaagatgtaactgtc 2268
Db 520 TGAGGTGGGAGGATCACTTTGAACCCAGGAGACAGAGGTTCAGTGAACCCGAGATCACGCC 579
QY 2269 acgaagggccagagtgagcagggagagcgtgccacatttgcgaagtgcctat----- 2321
Db 580 ACACGCTCCAGCTGGGCAACAGAACAGACTGTCTTAAAAAATAACAAAAATAA 639
QY 2322 -----gtgagggcggggagggcgggtcccttagagataagagacaatcataagggagat 2376
Db 640 AAGTAGTTCTACAGTACCAGCAATTCATTTTCAAAAGATATAGAGCTAAAAAGGAAGA 699
QY 2377 atcagagaaaaatcgtaagggagagcagatggttgcgaagagaataggctgaccatcgaagg 2436
Db 700 AAAAAAAGTAGTGTGGGCTTTAAATACCTGTTCTCTTAAATGCTTCTTAGGAGTG 759
QY 2437 actgscagagaagcttca-----gaaacacactggacggctgggcacagtgcc 2483
Db 760 CTGGGGTTTATGTCATCATTTATCCCTTTTAAATGTTATTTGGCCAGGCACGGTGGC 819
QY 2484 ttaggcctgtaatccagcactttgggaggtgcagcaggtgaaatcacttgaggtcagga 2543
Db 820 TCATGGCTGTAATCCAGCACTTTGGGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 879
QY 2544 gttccagaccagcctggccacacatggtgaaaccccatctctcagagaaaaataaaaaatta 2603
Db 880 GTGTGAGACCAACCGCTGGCCACATGGCGAAACCTG-TCTCTACTAAAAATAACAAAAATTA 938
QY 2604 gccagggcgtggtggcacaagcctagaatccacagctacttggagggcgtgagg 2654
Db 939 ACTAGGCGTGTGTGTACGCTGTAGTCCAGCTACTCGGGAGGCTGAGG 989
```

RESULT 11

us-08-487-002-20

Sequence 20, Application US/08487002

Patent No. 5710001

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Simard, Jacques

APPLICANT: Emi, Mitsuru

APPLICANT: Nakamura, Yusuke

APPLICANT: Durocher, Francine

TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer

TITLE OF INVENTION: Susceptibility Gene

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

STREET: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

```
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT: PatentIn Release #1.0, Version #1.30
APPLICANT NUMBER: US/08/487,002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-20
```

Query Match 6.2%; Score 165.4; DB 1; Length 6769;
Best Local Similarity 55.8%; Pred. No. 3.4e-24;
Matches 464; Conservative 0; Mismatches 321; Indels 46; Gaps 6;

```
QY 1869 aaaaattagctggcgtgtgtggcagatggttcagtagccgagacgcccactgca 1988
Db 160 AAAAAATTAGCTGGGTGTGGCAGCTACCTGTAAATCCAGCTACTCGGGAGCGGAGGC 219
QY 1929 atgagatcgtctgaacgggagagcagatgttcagtagccgagacgcccactgca 1988
Db 220 ACAAGAAATTCCTTGAACCTTAGCAGCGGGAGGTTCAGAGCGCAAGATCGCCACTGCA 279
QY 1989 ctccagcctggactacagagcagactctatctcaaaaaaaataaaaaaagt--a 2046
Db 280 CTCCAGCCTGGCCGTAGAGTGAAGTCTGTCTCAAAAAGAAAAAGTAATGTCTTCT 339
QY 2047 acttagtgaggggtgcctctgttattcaactgagacogtgcctccggttatgaggt--- 2103
Db 340 AGCTGGCGCAGTGGCTCTTGGCTGTAAATCCAGCACTTTGGGAGGCCAAGCGGTGGA 399
QY 2104 -----gtaccagaagaagatttcaactatgcacactattcacgcgtcacctagcatt 2157
Db 400 TCTCGAGTCTTGAAGTTCAGACCGCTAGGCAATGTGGTGAACCCCATCGCTACAAA 459
QY 2158 gaagccagcctgtagcctgaagaagccttctgttggagggcaggtcttctccc-----c 2208
```


THIS PAGE BLANK (USPTO)

Parent Case # 101767341

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 17:09:37 ; Search time 814.02 Seconds
(without alignments)
2795.191 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tgcctaggccatgctggc.....agctactggagggtgagg 2654

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	204	7.7	2713	20	AA27233
2	197.6	7.4	3621	22	AA500824
3	196	7.4	32204	22	AA157790
4	195.6	7.4	32152	22	AA157791
5	192.4	7.2	2498	22	AA29453
6	189.8	7.2	160552	22	AA02697
7	189.6	7.1	32199	22	AA157673
8	189.4	7.1	2839	22	AA38992
9	189.4	7.1	2839	22	AAH74642
10	189.4	7.1	2839	22	AA02498
11	186	7.0	41684	21	AAA28150

c 12	185.8	7.0	6711	21	AAF21122	Human low adenosin
c 13	185.8	7.0	6711	21	AAA35000	Human adenosine re
c 14	185.8	7.0	10032	21	AAF21124	Human low adenosin
c 15	185.8	7.0	10032	21	AAA35002	Human adenosine re
c 16	184.2	6.9	2446	22	AAH18305	Human cDNA sequenc
c 17	182.2	6.9	121162	21	AAH66548	Human kinesin-like
c 18	181.6	6.8	56583	21	AAF21125	Human low adenosin
c 19	181.6	6.8	56583	21	AAA35003	Human adenosine re
c 20	181	6.8	43069	21	AAZ36335	Genomic sequence o
c 21	180.4	6.8	23187	21	AA50273	Human lipolysis st
c 22	180.4	6.8	23187	22	AAE62331	Human leptin fragm
c 23	180	6.8	2001	22	AAH18441	Human cDNA sequenc
c 24	180	6.8	160552	22	AA02697	Human glycosyl sul
c 25	179	6.7	22976	20	AAH83426	Genomic region con
c 26	178.2	6.7	31853	22	AA163343	Human kidney relat
c 27	177.8	6.7	17431	22	AA157710	Human colorectal c
c 28	177	6.7	11967	22	AAE97853	Human neuroblastom
c 29	177	6.7	21721	20	AAH83427	Human lipolysis st
c 30	176.8	6.7	13927	22	AAE97853	Human neuroblastom
c 31	176.8	6.7	57728	22	AAE87588	Human 9p11 chromos
c 32	176.8	6.7	121162	21	AAH66548	Human kinesin-like
c 33	176	6.6	84607	20	AAE90847	Human PACAP genomi
c 34	174.2	6.6	78925	21	AAE89888	Human FN gene. Ho
c 35	174	6.6	6405	22	AAE97850	Human neuroblastom
c 36	173.2	6.5	110000	22	AAE84800	Nucleotide sequenc
c 37	173	6.5	14426	22	AA162921	Human genomic DNA
c 38	173	6.5	26928	20	AAZ32184	Human prothrombin
c 39	172.6	6.5	32134	22	AA163522	Human kidney relat
c 40	172.6	6.5	32192	22	AA163523	Human kidney relat
c 41	172.2	6.5	32249	22	AA162932	Human genomic DNA
c 42	172.2	6.5	53526	19	AAE94101	Human PKD1 gene.
c 43	172.2	6.5	53577	17	AAE18551	Human polycystic k
c 44	172.2	6.5	53577	19	AAE94108	Human PKD1 locus b
c 45	171.4	6.5	5197	22	AA162920	Human genomic DNA

ALIGNMENTS

RESULT 1
AAZ27233
ID AAZ27233 standard; DNA; 2713 BP.
XX.
AC AAZ27233;
XX
DT 28-MAY-1999 (first entry)
XX
DE Human Rab protein, RABP-3, coding sequence...

XX Rab protein, RABP-1; RABP-2; RABP-3; human; vesicle trafficking; cancer;
XX cell differentiation; apoptosis; immunodeficiency; cell proliferation;
XX neurodegenerative disease; myelodysplastic syndrome; wasting disease;
XX toxin-induced disease; infection; genetic defect; diagnosis; therapy; ss.

OS Homo sapiens.

PN WO9909182-A2.

XX 25-FEB-1999.

PD 17-AUG-1998; 98WO-US16983.

XX 21-AUG-1997; 97US-0916901.

XX (INCY-) INCYTE PHARM INC.

XX Corley NC, Hillman JL, Lal P, Shah P;

XX WPI; 1999-181042/15.

XX P-PSDB; AAY00920.

XX New purified human Rab proteins - used to develop products for

XX treating e.g. AIDS, immunodeficiencies, neurodegenerative diseases,

Qy 1970 ccgagacggccactgcactccagcctgggactacagagcagagactctatctcaaaaaa 2029
Dy 23894 CCAAGATCGACCACTACACTCCAGCATGGGCAACACAGAGAGAGCTGTCTCAAAAAA 23835
Qy 2030 aaaaaa-aaagtaactttagtgaggggtgctctgttatttactcagagaccgtg 2087
Dy 23834 TAAATAAATAGGCCAGGTGCGAGTGGCTCATGCTGTAATTCAGCAGCATTTGGGAGGCCGAG 23775
Qy 2088 ccccggttatgaggtgtaccagaaaagcaagatttaccatgtacacactatcacgcgtca 2147
Dy 23774 CGAGCCAGATCAGAGTCTAGGAGTTCAGACACGCCCTGGCCAAATATGGTGACACCCCC 23715
Qy 2148 ccttagcattgaagccagcctgtagcctgaaagcccttg-----ctttgagggcaggt 2200
Dy 23714 GTCTCTACTAAAAATACAAAAATAGCTGGCGGTGGTGGCTCGCGCCTTTAGTCCCACT 23655
Qy 2201 ctttcccaaaatcagacacgaaggtgcaaaagtgaagtcctcagtccttgcacaaagtgt 2260
Dy 23654 ACTTGGGAAGCTGAGGCAAGAAATCGCTTGAACCCAGGAGCGGAGATTTTCAGTGAGCC 23595
Qy 2261 aactgtcacgaagccacagagtgagcagggagagctgtccccacatttgcggaagtggcta 2320
Dy 23594 AAGATGGTCCCACTG-CACTCCAGCCTGGTGACAGACCAAGACTCTCTCAAAAAATA 23536
Qy 2321 tgtgagggacggggagggcgggtcccttagagataagagacacataagaggggagatatca 2380
Dy 23535 AATAAATAAATAAATAATGTGCCAGCTGTGGCTCACCCTTATAATCCTAGCACCTTTGGG 23476
Qy 2381 gaaaaaatcgaagggagagagatggtgtcaagagaataagctgaccatgaaggaactg 2440
Dy 23475 AAGTTGA--GGTGGGTGGATTGCTGTGATCTCAGGATTAACAGCACGAGCTAGGCAACATAG 23418
Qy 2441 gcagaagcttcc---agaaaaccactggagcggctgggcacagtggtccttagcctgtaact 2497
Dy 23417 TGAGACCTCATCTCAATAAATCAATAAATAGCTGGGCGCAGTGGCTCATGCCGTGTAATC 23358
Qy 2498 ccagcactttggaggctgacgcagggtgaatcaccttaggtcagggtccagagccagcc 2557
Dy 23357 CCAGCACTTTGGAGGCTGAGGTGGCGGATCACTTGAGGTGAGGTTCCAGAGCCAGCC 23298
Qy 2558 tggcaacatggtgaaccccatctctacagaaaataaaaaattagccagggctggtgg 2617
Dy 23297 TTGCCAATGATGAACCTTGTCTCTACTAAATAACAAAAATAGCTGGGCGCAGGTGG 23238
Qy 2618 cacaagctagaatccagctacttgggaggtgag 2654
Dy 23237 CACACACCTGTAGTCCCACTATTTGGGGCCTGAGG 23201

RESULT 3
AAI57790
ID AAI57790 standard; DNA; 32204 BP.
XX XX
AC AAI57790;
XX XX
DT 19-OCT-2001 (first entry)
XX XX
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 327.
XX XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX XX
OS Homo sapiens.
XX XX
PN WO200155350-A1.
XX XX
PD 02-AUG-2001.
XX XX
PF 17-JAN-2001; 2001WO-US01350.
XX XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR

01-NOV-2000: 2000US-0244617.
08-NOV-2000: 2000US-0246474.
08-NOV-2000: 2000US-0246475.
08-NOV-2000: 2000US-0246476.
08-NOV-2000: 2000US-0246477.
08-NOV-2000: 2000US-0246478.
08-NOV-2000: 2000US-0246523.
08-NOV-2000: 2000US-0246524.
08-NOV-2000: 2000US-0246525.
08-NOV-2000: 2000US-0246526.
08-NOV-2000: 2000US-0246527.
08-NOV-2000: 2000US-0246528.
08-NOV-2000: 2000US-0246532.
08-NOV-2000: 2000US-0246609.
08-NOV-2000: 2000US-0246610.
08-NOV-2000: 2000US-0246611.
08-NOV-2000: 2000US-0246613.
17-NOV-2000: 2000US-0249207.
17-NOV-2000: 2000US-0249208.
17-NOV-2000: 2000US-0249209.
17-NOV-2000: 2000US-0249210.
17-NOV-2000: 2000US-0249211.
17-NOV-2000: 2000US-0249212.
17-NOV-2000: 2000US-0249213.
17-NOV-2000: 2000US-0249214.
17-NOV-2000: 2000US-0249215.
17-NOV-2000: 2000US-0249216.
17-NOV-2000: 2000US-0249217.
17-NOV-2000: 2000US-0249218.
17-NOV-2000: 2000US-0249244.
17-NOV-2000: 2000US-0249245.
17-NOV-2000: 2000US-0249264.
17-NOV-2000: 2000US-0249265.
17-NOV-2000: 2000US-0249297.
17-NOV-2000: 2000US-0249299.
17-NOV-2000: 2000US-0249300.
01-DEC-2000: 2000US-0250160.
01-DEC-2000: 2000US-0250391.
05-DEC-2000: 2000US-0251030.
05-DEC-2000: 2000US-0251988.
05-DEC-2000: 2000US-0256719.
08-DEC-2000: 2000US-0251479.
08-DEC-2000: 2000US-0251856.
08-DEC-2000: 2000US-0251868.
08-DEC-2000: 2000US-0251869.
08-DEC-2000: 2000US-0251989.
08-DEC-2000: 2000US-0251990.
11-DEC-2000: 2000US-0254097.
05-JAN-2001: 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457727/49.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the colon and rectum including colorectal cancers and also for testing and detection e.g. diagnosis -

Disclosure; SEQ ID NO: 328; 522pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of colorectal cancer antigens. These are shown in AA157547-AA157619 and AA385569-AA386641. These can be used in the diagnosis, prevention and treatment of cancer of the colon and/or rectum. The present sequence is a colorectal cancer antigen genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pt_sequences.

Sequence 32152 BP; 8332 A; 7335 C; 7256 G; 9229 T; 0 other;

Query Match 7.4%; Score 195.6; DB 22; Length 32152;
Best Local Similarity 56.5%; Pred. No. 6.9e-30;
Matches 446; Conservative 0; Mismatches 334; Indels 10; Gaps 4;
QY 1869 aaaaattagctggcgctggatgcatccacaatacccagctactctgggagggctgagggc 1928
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8688 aaaaattagctggcgctggatgcatccacaatacccagctactctgggagggctgagggc 8747
QY 1929 atgagaatcgttgaaccggggagggagagatgttcagtgagccgagcgcgcacactgca 1988
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8748 aggagaatggcttgagcccaggggagggaggttcagtaagctgagatgatgctgcaactgca 8807
QY 1989 ctccagcctggactacagagcagactctctcaaaaaaaataaaaaaaagtaaac 2048
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8808 ctccagcctggagacagagtgagactctctcaaaaaaaataaaaaaaagtaaac 8867
QY 2049 ttaggctcaggggtgctctctgttattcactgaga---ccgtgcccgggttatgagggttgt 2105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8868 aatatcagatggccaggccaggcagtggtgctcacactgtaatacacgactatggaagg 8927
QY 2106 accagaaagcaagtattcactatcgcacactattccacgctcaccttagcattgagccag 2165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8928 ctgaggcaggaggttgcttaagcccttagaggttcaagaccagactgggccaatgtagtgg 8987
QY 2166 cctgtagcctgaaagccttctgttggggcaggtcttcccccaaaatgcagacacgaag 2225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8988 ----accccccatctctacaggaacaataaaatacaataaaatggcagttcttcag 9042
QY 2226 gtcaaaagtgaagctgcagctcttgcaaaagatgtaactgtcac-gaaggccacgagtg 2284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9043 ttgtctgtctgggttctctatgtgtcactgtctcttcagaaatcactgtatggccaggttca 9102
QY 2285 gcaggggagactgtcccaatttcggaagtgtgtatgtgagggcggggagggcggtgcc 2344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9103 ctcttcagactggaaagcagcagcacaaggtatgggtatttcaggggccagaaactccaggt 9162
QY 2345 cttagagataagagacataaaggggagagatatcagagaaatcgttaaggggagcgagat 2404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9163 atttttaatttttttaatttaaaattatacaggggccaaagcag-gtggctcacacct 9221
QY 2405 ggttgtaagagaataggttgacctgacactgagaggactgagaaagtcttcagaaacccactg 2464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9222 gtaattccggtactttggagaaacatgcaaaacccctgtctctactaaaaatacaaaaa 9281
QY 2465 gacgctgggacacagtgcttagccttagccttaatcccagcactttgggaggtgagcgaggt 2524
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9282 tcaggctggcgcggttagctcatcctgttaatcccagcactttgggagggccaaaggcaggt 9341
QY 2525 gaatcacttgaggtcagaggttcacagaccagcctggccaacatggtgaaaccccatctct 2584
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9342 ggatcacctgagatcaggagctgagccagcctggccaacatggtgaaaccccatctct 9401
QY 2585 acagaaaaataaaaaattagccaggcgtgtgtggcacaaagcctagaatcccagctacttgg 2644
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9402 actaaaatacaaaaattagccgggcatggtgtgtacacacttttaatacccagctactaaag 9461
QY 2645 gaggtgagg 2654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9462 gaggtgagg 9471

RESULT 5
AAF29453
ID AAF29453 standard; cDNA; 2498 BP.
XX
AC AAF29453;
XX
DT 03-APR-2001 (first entry)
XX
DE Human TANGO 292 cDNA.
XX
KW Human; INTERCEPT 217; INTERCEPT 297; TANGO 276; TANGO 292; TANGO 325;

RESULT	6
AAD02697	
ID	AAD02697 standard; DNA; 160552 BP.
XX	
AC	AAD02697;
XX	
DT	02-MAY-2001 (first entry)
XX	
XX	Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
XX	
KW	Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demylinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW	chromosome 16q23.1; ds.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
exon	32847..32922
FT	/*tag= a
FT	/number= 1
FT	/label= 4a_504
intron	32923..35592
FT	/*tag= b
FT	/cons_splice= (5'site:NO, 3'site:YES)
exon	35593..35674
FT	/*tag= c
FT	/number= 2
FT	/label= 4a_503
intron	35675..45093
FT	/*tag= d
FT	/number= 3
exon	45094..45185
FT	/*tag= e
FT	/number= 4
FT	/label= 4a_502
intron	45186..46633
FT	/*tag= f
FT	/cons_splice= (5'site:NO, 3'site:NO)
exon	46634..46700
FT	/*tag= g
FT	/number= 4
FT	/label= 4a_501
intron	46701..47938
FT	/*tag= h
FT	/cons_splice= (5'site:YES, 3'site:NO)
exon	47939..47955
FT	/*tag= i
FT	/number= 5
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF and all of 3'UTR"
5'UTR	47939..47955
FT	/*tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
CDS	47956..49128
FT	/*tag= k
FT	/product= "Human glycosyl transferase-4alpha (GST-4alpha)"
3'UTR	49129..49746
FT	/*tag= l
exon	83257..83347
FT	/*tag= m
FT	/label= 4a_502
intron	83348..96412
FT	/*tag= n
FT	/cons_splice= (5'site:NO, 3'site:NO)
exon	96413..96484
FT	

```
Qy 1989 ctccagcctgactacagacgagcactctctcaaaaaaaaaaaaaaaaaaagttaac 2048
|||||
Db 55111 ctccagcctggcgaatagctgagcactctctcaaaaaaaaaaaaaaaaaaagttaac 55170
|||||
Qy 2049 ttaggtcagaggtgctctctgttattcaactgagacgctgccccggttatgaggtgttacc 2108
|||||
Db 55171 gactatgagcttgaagagacacaaactc-----tttgaaca 55208
|||||
Qy 2109 agaagaagcatttcaactatgcacactattccacgctcaccctagcattgaagccagcct 2168
|||||
Db 55209 agacacaaatcccaaaactgtgagaagaattcactttcaaaataataataaaggagc 55268
|||||
Qy 2169 gtacgtgaagccttctgtcttgaggcagctcttcccaaaatgcagacacgaaggtg 2228
|||||
Db 55269 aaaacttttgaattgctgtgtgtgacgtggtgatctgctcagaggtctctatactc 55328
|||||
Qy 2229 caaagtga-agctgccagcttgcacaaagatgtaactgtcacgaagccacgagtgga 2287
|||||
Db 55329 taattttacacttacatataaaattttataataaattatgtctcaatactttttttac 55388
|||||
Qy 2288 gggagagctgtccacatttgcgaagtggctatgtgaggacggggagcggttccctt 2347
|||||
Db 55389 agaagcattggggaaataaccataaagtaggttaaaaaacagatg-----a 55435
|||||
Qy 2348 agagataagagacacataaagggagatatacagagaaatcgtaaagggagcagatggt 2407
|||||
Db 55436 caagctagcagaacacatttgcataaaaaataagacaaaacccagaggaagcgaagt 55495
|||||
Qy 2408 tgtcaagaagatagctgacatcgaaggactgagcagaagctttcagaaaaaccactggac 2467
|||||
Db 55496 caacagaaaaatgagcaagccctttacaag-----aaaaaagaagccca 55542
|||||
Qy 2468 gctgggcacagctgcttagcctgttaatccacagcactttggggagctgaagcaggtgaa 2527
|||||
Db 55543 gggcggcgcagtggtccacgcctgtaatccacagcactttggggagctgaggtgggtgga 55602
|||||
Qy 2528 tcaactgaggtcagaggttccagaccagcctggcacaactgtgtaaaccccatctctaca 2587
|||||
Db 55603 tcaactgaggtcagaggttggagaccagcctggcacaactggcacaaccccgctctact 55662
|||||
Qy 2588 gaaaataaaaaatagccagcgtggtggcacaagcctagaatccacagctacttgggag 2647
|||||
Db 55663 aaaaataaaaaatagccggtgtatggcacacgcctgtagcccgacttacttcagaa 55722
|||||
Qy 2648 gctgagg 2654
|||||
Db 55723 gctgagg 55729

RESULT 7
AAI57673
ID AAI57673 standard; DNA; 32199 BP.
XX
AC AAI57673;
XX
XX 19-OCT-2001 (first entry)
XX
XX Human colorectal cancer antigen coding sequence SEQ ID NO: 210.
XX
XX Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200153350-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01350.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
```

```
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
```


02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-457727/49.
 XX
 XX Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the colon and rectum including colorectal cancers
 PT and also for testing and detection e.g. diagnosis -
 XX
 XX Disclosure; SEQ ID NO: 210; 522pp + Sequence Listing; English.
 PS
 XX


```
FT /*tag= x
FT /number= 12
FT intron 28829..34698
FT /*tag= y
FT /number= 12
FT exon 34699..34791
FT /*tag= z
FT /number= 13
FT intron 34792..36678
FT /*tag= aa
FT /number= 13
FT exon 36679..36361
FT /*tag= ab
FT /number= 14
FT intron 36862..39013
FT /*tag= ac
FT /number= 14
FT exon 39014..39169
FT /*tag= ad
FT /number= 15
FT intron 39170..39455
FT /*tag= ae
FT /number= 15
FT exon 39456..39684
FT /*tag= af
FT /number= 16
FT variation replace(6491,T)
FT /*tag= ag
FT variation replace(15234,G)
FT /*tag= ah
FT variation replace(15868,T)
FT /*tag= ai
FT variation replace(16729,T)
FT /*tag= aj
FT variation replace(18311,C)
FT /*tag= ak
FT variation replace(18572,T)
FT /*tag= al
FT variation replace(22906,A)
FT /*tag= am
FT variation replace(23175,T)
FT /*tag= an
FT variation replace(23253,G)
FT /*tag= ao
FT variation replace(26106,C)
FT /*tag= ap
FT variation replace(30464,G)
FT /*tag= aq
FT variation replace(30669,C)
FT /*tag= ar
FT variation replace(31250,A)
FT /*tag= as
FT variation replace(35148,G)
FT /*tag= at
FT variation replace(36801,G)
FT /*tag= au
FT variation replace(37286,T)
FT /*tag= av
FT variation replace(37536,C)
FT /*tag= aw
FT variation replace(39321,A)
FT /*tag= ax
FT variation replace(39689,T)
FT /*tag= ay
FT primer_bind 6137..6157
FT /*tag= az
FT primer_bind complement (6577..6597)
FT /*tag= ba
FT primer_bind 14864..14882
FT /*tag= bb
FT primer_bind complement (15292..15312)
FT /*tag= bc
FT primer_bind 15837..15855
```

```
FT /*tag= bd
FT complement (16429..16266)
FT /*tag= be
FT primer_bind 16599..16617
FT /*tag= bf
FT complement (17030..17049)
FT /*tag= bg
FT primer_bind 18131..18150
FT /*tag= bh
FT complement (18592..18610)
FT /*tag= bi
FT 22710..22727
FT /*tag= bj
FT 23100..23118
FT /*tag= bk
FT complement (23130..23149)
FT /*tag= bl
FT complement (23512..23530)
FT /*tag= bm
FT 25822..25840
FT /*tag= bn
FT complement (26222..26241)
FT /*tag= bo
FT 30332..30352
FT /*tag= bp
FT complement (30754..30773)
FT /*tag= bq
FT 30918..30935
FT /*tag= br
FT complement (31390..31408)
FT /*tag= bs
FT 34780..34799
FT /*tag= bt
FT complement (35215..35233)
FT /*tag= bu
FT 36593..36610
FT /*tag= bv
FT complement (37017..37036)
FT /*tag= bw
FT 37060..37080
FT /*tag= bx
FT complement (37541..37561)
FT /*tag= by

Query Match 7.0% Score 186; DB 21; Length 41684;
Best Local Similarity 56.8% Pred. No. 5.7e-28;
Matches 452; Conservative 0; Mismatches 320; Indels 24; Gaps 5;

Qy 1869 aaaaattagctggcggtggtgcatccacatccacagctactggtggaggtgagggc 1928
Db 7592 AAAAAATTAGCGGAGCGTGTGCGGGCACCTGTATCCAGCTACTGGGAGGCTGAGGC 7533
Qy 1929 atgagaatcgcttgaaccgggggagcgagatgttcagtgagccgagagcgccactgca 1988
Db 7532 AGGAGAATACACTTGAACCCCAAGAGTTGGAGGTGCGAGTCGAGGATCATGCCACTGCA 7473
Qy 1989 ctccagcctgactacagagcgagactctatctcaaaaaaataaaaaa-----aa 2039
Db 7472 CTCAGCGCTGGCGGACAGAGCAAGACTCCATCTCAAAAAACAAACAAACACATGAA 7413
Qy 2040 aaagtaacttagtgaggggtgctctctgttattcactgagagccgtgcccgcgttata 2099
Db 7412 GATGCATATCAAGTGAAGACATAAAAGACTGTTTAAATAAGTGGTGTGCTATTG 7353
Qy 2100 ggtgtaccagaagaagtaattcactatgcacactatccaccgctccaccctagcattga 2159
Db 7352 GAATAACGGGTTAAACCCCAAAATAATATTCCTTATGAATCAAAAGGTTTAGCATTTA 7293
Qy 2160 agccagcctgagcctaagcccttctgtgagggcaggtctttcccaaaatgcacac 2219
Db 7292 AAAAATCATGCATGTAGCAA-AAITTAATAATGTGAGGAATGCTTTTCCAAGTTTAAACAA 7234
Qy 2220 acgaaggtgcaagtgaaagtcgagtcgagtccttgcataaagatgttaacttgtcaagagggccac 2279
```

Db	7233	AAGATTGAGAGTTCGATTACAGAAATGTTAAATAAAAAGTTTACATGGGAAAAAAAAG	7174
Qy	2280	gagtggcaggagagctgccccacattgcggaagtggctatgtagagacggggagcg	2339
Db	7173	TAAACTAGAAAAAATAATCACACATGTGCANAAAGGTCAAGTGAAGGACAGTAAATATT	7114
Qy	2340	ggtcctctagagataagagacataataagggagatcatcagagaaatcgtaaagggag	2399
Db	7113	AATCCATATACAAAGCAGACCCCTAGAAAGTCATAACCGAACACTCTGGCA-----	7058
Qy	2400	cagatggttgtcaagagaatagctgaccatcgaaagactggcagaagcttcctcagaataac	2459
Db	7059	-----GACTTTATAAGTTGTAATGGTGTGTAATAATGCCCACAAATATATAAGAAAAAAT	7006
Qy	2460	cactggacggctggccacagtagcttagccctgtaatccceagacactltggagagctgacg	2519
Db	7005	AAAGCAGGGGCCACAGTGGGGTGGCTCACACTGTAAATCCCAGCACTTTGGAGAGCCAAAG	6946
Qy	2520	caggtgaactcacttgaggtca-ggagttccagaccagcctgtgcaacatggtgaacccc	2578
Db	6945	CGAATGCATCACTGAGGTAAAGGGAGTTCACACCGCCTGACCAACATGCTGNAACCCC	6886
Qy	2579	atctctacgagaataataaaaaattgccaggcgtggtggccaagcctgaataccccagct	2638
Db	6885	GCCTCTCTAAAAATA-CAAAATTAGCCAGGCATGGTGGGCATGCCTATAATCCCCAGCT	6827
Qy	2639	acttggagagctgagg	2654
Db	6826	ACTTGGGAGGCTGAGG	6811

RESULT 12

AAF21122/c
ID AAF21122 standard; DNA; 6711 BP.

AA	AAF21122;
AC	
XX	14-MAR-2001 (first entry)
XX	
DT	
XX	
DE	Human low adenosine antisense oligonucleotide related sequence #2689
XX	
XX	Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW	human; airway disorder; bronchoconstriction; lung inflammation;
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW	immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW	respiratory obstruction; pulmonary obstruction; impeded respiration;
KW	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis
KW	pulmonary hypertension; emphysema; pulmonary transplantation rejected
KW	chronic obstructive pulmonary disease; pulmonary infection; bronchitis
KW	cancer; ss.

XX
OS Homo sapiens.

XX
PN WO200062736-A2.

XX
PD 26-OCT-2000.

XX
PF
24-MAR-2000; 2000WO-US08020.

XX
PR 06-APR-1999; 99US-0127958.

XX
PA (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.
XX

PI Nyce JW;
XX

DR	WPI; 2000-679539/66.
XX	
XX	
PT	Low adenosine (A) content antisense oligonucleotides which do not
PT	trigger adenosine receptors during metabolism, useful e.g. for treating
PT	cancers and respiratory obstructions -

[illegible]

Query Match 7.0%; Score 185.8; DB 21; Length 6711;
Best Local Similarity 55.7%; Pred. No. 4.8e-28;
Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;

QY	1869	aaaaattagctggcgctgtggcatgcatccaaatccccagctactctgaggcgctgagcc	1828
Db	5916	AAAAATTAGCCGGCATGTGGCGATTGCCTATAGTCCAGCTACTCGGAGGCTGAGCC	5857
QY	1929	atgagaatcgcttgaaccgggagcgagatgttcagtgagccgagcgcgccactgcga	1988
Db	5856	AGGAGAAATGGCGTGAACCTTGGGAGCGGAGCTTGCAGTGACCCGAGATCGGCCACTGCA	5797
QY	1989	ctccagcctggactacagagcgagactcttatctcaaaaa-aaaaaaaaaaaaaaagtaa	2047
Db	5796	CTCAGCCTGGCGCAGCAGACGAAACTCCATCTCAAAAAGAGTTACACATATAGAATGA	5737
QY	2048	cttagctgagggctgcctctgtattcactgagaccgtgcccggttatgaggtgtgac	2107
Db	5736	ATTGGCCCCCTCCCCAAAAACAACAAAACAACCAAAAAACCCCAAGAACAAAAA	5677
QY	2108	cagaagcgaagtattactactgacactattccacgcgtcacccctagcattgaagccagcc	2167
Db	5676	ACCCAACCAACCGCATGTTATTATAGTAACATAAAAGACGACCTTAATTTGGTGGGCAC	5617
QY	2168	tgtagcctgaagccttctgtcttgaggcgaggtcttctcccaaatcagacacgcaaggt	2227
Db	5616	AGTGGCTCATGCGCTGTAAATCCAGACACTTGGGGGGGTCAAGGCAGCGACATCACTTGAG	5557
QY	2228	gcaaatgtaagctgccagctcttgcaaaaagatgaacttgtc-----acgaagggccac	2279
Db	5556	CCCAGGAGTTCGAGACACGCGCTGGACAACATGGTAAATCCCGCTCTACTAATAATACACA	5497
QY	2280	gagtgcaagggagagctgtcccaactttgcggaagtgacctatgtagagacgggggagcgc	2339
Db	5496	ATTTAGCTGGGAGTGGTGATGCCCTGTAGTTCAGCTACTTTGAG--AGGCTGAGGCA	5439
QY	2340	ggtcccttagagataagagacaaatcataagggagatcatcagaaataatcgtaaagggag	2399

Db 5438 GAAGAATTGCTTGAACCCGGGAGGAGGTTTCAGTGTAGCTGAGTGGCCACACTGCCTCC 5379
QY 2400 cagatgttctcaagagaaatagctgaccatcgaaagactggcagaagctttcagaaaac 2459
Db 5378 AACTGGGACATAGAG-----CGAGATCCAGTCTGAGAAAAGAAAAAAGG 5323
QY 2460 cactggacgctgggcacagcttagcttagcctgtaatccccagcactttggagggctgacg 2519
Db 5322 ACTTAATTGTTGGCGCAGTGTGCTCATGCTGAAATCCAGCAGCTTTGGGGGCCAAGG 5263
QY 2520 cagtgtaacacttgagctcagagtggtccagacagcagcctggcacaatggtgaaacccca 2579
Db 5262 CAGGTGTTTACCTTAAGTACAGAGTTCAAGACCAAGCTTGGCCAAACATGATGAACCCCG 5203
QY 2580 tctctacagaaaataaaaaattagccagcggtggtggcacaagccttagatccccagcta 2639
Db 5202 CTCACAAAAATACAAAAAATTAGCCAGGCGATGCTGGCCACACATCTGTATTCTCAGCTA 5143
QY 2640 cttggggagggctgag 2654
Db 5142 CTCAGGAGGCCAAGG 5128
RESULT 13
ID AAA35000/c
AC AAA35000;
XX 28-JUL-2000 (first entry)
XX Human adenosine receptor related polynucleotide SEQ ID NO:2689.
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antisthmatic; cytosolic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX Homo sapiens.
XX OS
XX W0200009525-A2.
XX 24-FEB-2000.
XX 03-AUG-1999; 99WO-US17712.
XX 03-AUG-1998; 98US-0095212.
XX (UYEC-) UNIV EAST CAROLINA.
XX PA
XX NYce JW;
XX WPI; 2000-205971/18.
XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension, or
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
XX PS Disclosure; Page 898-900; 1343pp; English.
XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antisthmatic, cytosolic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary

CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA3512 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1880
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 6711 BP; 1552 A; 1625 C; 1717 G; 1817 T; 0 other;
Query Match 7.0%; Score 185.8; DB 21; Length 6711;
Best Local Similarity 55.7%; Pred. No. 4.8e-28;
Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;
QY 1869 aaaaattagctggcgctggtggcatgcatacccaatcccagctactgggagggctgagggc 1928
Db 5916 AAAAAATTAGCGGGCATGTTGGCATGTGCTTATAGTCCAGCTACTCGGAGGCTGAGGC 5857
QY 1929 atgagaatcgtttgaaccgggggagcgagatgttgcagtgcgcgcgcgcgcgcgcgcgcgc 1988
Db 5956 AGGAGAATGCGTGAACCTGGGAGCGGAGCTTGCAGTGCAGCCGAGATCGGCCACTGCA 5797
QY 1989 ctccagcctggactacagagcagagactctctcaaaaa-aaaaaataaaagtaa 2047
Db 5796 CTCAGGCTGGGCGACAGAGCGAACTCCATCTCAAAAAGAGTTACACATATAGAATGA 5737
QY 2048 cttagtgaggggtgctctgttattcactgcagacgcgtgcccgttatgaggtgtgac 2107
Db 5736 ATTGCGGCCCTCCCAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 5677
QY 2108 cagaaagcaagatttaccactgcacactattccgcgtccaccctagcattgaagccagcc 2167
Db 5676 ACCCAACCAACCGCATGTTATTAATAGTAAACAAAGCAGCAGCTTAATTGGTGGGCAC 5617
QY 2168 ttagcctgaaagccttctgttgagggcgaggtcttcccccaaaatgcagacacgaaggt 2227
Db 5616 AGTGGCTCATGCTGTAATCCAGCAGCTTTCGGGGGGGTCAAGGAGGAGATCATCTTGAG 5557
QY 2228 gcaaaagtgaagctgcagctcttgcacaaagatgtaactgtc-----acgaagggccac 2279
Db 5556 CCCAGGAGTTGCGAGACCGAGCTGGACACACATGGTAAATFCCCGCTCTACTAATAATACACA 5497
QY 2280 gagtggcagggagagctgtccacatttgcggaagtggtgagcagcggggagggcg 2339
Db 5496 ATTTAGCTGGGAGTGGTGGTGCATGCCCTGTAGTTCAGCTACTTTGAG--AGCTGAGGCA 5439
QY 2340 ggtcccttagagaataagagacaatcataaggggagatatcagagaaaatcgtgaagggag 2399
Db 5438 GAACAATTGCTTGAACCCGGGAGGAGGTTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 5379
QY 2400 cagatggttctcaagagaaatagctgaccatcgaaagactggcagaagctttcagaaaac 2459
Db 5378 AACCTGGGAGATAAGAG----CGAGATCCAGCTCTGAGAAAAGAAAAAAGG 5323
QY 2460 cactggacgctgggcacagcttagccttagcctgtaatccccagcactttggagggctgacg 2519
Db 5322 ACTTAATTGTTGGCGCAGTGTGCTCATGCTGAAATCCAGCAGCTTTGGGGGCCAAGG 5263
QY 2520 cagtgtaacacttgagctcagagtggttccagacagcagcctggcacaatggtgaaacccca 2579
Db 5262 CAGGTGTTTACCTTAAGTACAGAGTTCAAGACCAAGCTTGGCCACACATCTGTATTCTCAGCTA 5203

RESULT 15

AAA35002/c
ID AAA35002 standard; DNA; 10032 BP.

AC AAA35002;

28-JUL-2000 (first entry)

Human adenosine receptor related polynucleotide SEQ ID NO:2691.

Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytosstatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX W0200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US1712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 900-903; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytosstatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 10032 BP; 2342 A; 2486 C; 2643 G; 2561 T; 0 other;

Query Match 7.0%; Score 185.8; DB 21; Length 10032;
Best Local Similarity 55.7%; Pred. No. 5.1e-28;
Matches 443; Conservative 0; Mismatches 337; Indels 15; Gaps 4;

Qy	1869	aaaaattagctggcggtggatgcatccacaatcccagctactgggagggctgaggc	1928
Db	8514	AAAAATTAGCGGGCATGGTGGCATGTATAGTCCAGCTACTCGGAGGCTGAGGC	8455
Qy	1929	atgagaatcgcttgaaacgggggagcgagatgttcagctgagccgagacgcgccactgca	1988
Db	8454	AGGAGATGCGTGAACCTGGGAGCGGAGCTTCAGTGAGCCGAGATCGGCCACTGCA	8395
Qy	1989	ctcagcctggactacagagcgagactctatctcaaaaa-aaaaaataaaataaagtaa	2047
Db	8394	CTCAGCCTGGGCGAGAGAGAACTCCATCTCAAAAAGAGTTACACATATAGAATGA	8335
Qy	2048	cttaggtgcaagggtgctctctgttattcaactgagaccgtgccccgggttatgaggtgtac	2107
Db	8334	ATTGCGCCCTCCCTCCCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	8275
Qy	2108	cagaaagcaagatttcactatgcacactattccacgctcacccctagcattgaagccagcc	2167
Db	8274	ACCAACCAACCGCATGTTATTATAGTAACATAAACAGCAGCTTAATTGGTGGGCAC	8215
Qy	2168	tgtagcctgaaagccttctgtttgagggcgaggtcttctcccaaaatgcagacacgaaggt	2227
Db	8214	AGTGGCTCATGCTGTAATCCAGCACTTCGGGGGGTCAAGGCAGGCAGATCATCTTGAG	8155
Qy	2228	gcaagtgaagctgcagctcttgcaaaagatgtaacttgtc-----acgaagggccac	2279
Db	8154	CCCAGAGATTTCAGAGACCAGCCTGGACCAACATGGTAAATCCCGCTCTACTAATAATACACA	8095
Qy	2280	gagtgccagggagagctgccacatcttcggaagtggtctatgtgagacggggagggcg	2339
Db	8094	ATTTAGCTGGGAGTGGTGGTCATGCCCTGTAGTTCAGCTACTTTGAG--AGGCTGAGGCA	8037
Qy	2340	ggtcccttagagataagagacaataaaggggagagatatcagagaaaaatcgtaaagggag	2399
Db	8036	GAAGAAATTGCTTGAACCCGGGAGGAGGTTCAGTGAGTGAGTGGCACCACTGCACATCC	7977
Qy	2400	cagatggttgcagagagaataggtgacctcgaaggaactggcgagagctttcagaaaaac	2459
Db	7976	AACCTGGGAGATAAGAG---CGAGATCCAGTCTCAGAAAAAGAAAAAAGG 7921	
Qy	2460	cactggagcgctggcgacagtggtcttaggcctgtaatcccagcactttggagggctgacg	2519
Db	7920	ACTTAATTGGTTGGCGCAGTGGCTCATGCTGAATCCAGCACTTTGGGGGGCCAAAGG	7861
Qy	2520	cagdtgaatcacttgaggtcagggaggttcagaccagcctggcacaacatggtgaacccca	2579
Db	7860	CAGGTGGTTACCTTAAGTCACGAGTTCAAGACCAGCTTGGCCAAACATGATGAACCCCG	7801
Qy	2580	tctctacagaaaaatataaaaaattagccagggcggtggtggcacaagccctagatccacagcta	2639
Db	7800	TCTCCACAAAAAATACAAAAAATAGCCAGGCATGCTGGCACACATCTGTATCTCAGCTA	7741
Qy	2640	cttgggaggtgagg 2654	
Db	7740	CTCAGGAGGCCAAGG 7726	

Search completed: January 11, 2002, 17:19:41
Job time: 29533 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:36:18 ; Search time 8779.43 Seconds
(without alignments)
3248.422 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tgcctagccatagcttggc.....agctacttggagagctgagg 2654

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estov:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	420.4	15.8	538	13	AQ223799 HS_2218_A
C 2	335.6	12.6	476	10	A1760998 wh97c08.x
C 3	332.6	12.5	478	10	A1540062 td08g08.x
C 4	332.6	12.5	594	10	AW515919 xy02c11.x
C 5	331.2	12.5	444	10	AW015376 UI-H-B10-
C 6	331.2	12.5	750	10	AW072388 xa07d05.x
C 7	330.4	12.4	441	10	A1239613 qh37d07.x
C 8	330	12.4	561	10	A1682248 wa71q04.x
C 9	329.4	12.4	451	10	AW572661 xx92b07.x
C 10	329.4	12.4	625	10	A1479089 tm31a08.x
C 11	318.2	12.0	337	10	AW977599 EST389708
C 12	287.4	10.8	311	10	AW207763 UI-H-B11-

C 13	281.4	10.6	547	10	A1436579
C 14	280.8	10.6	877	10	AA767386 nt81n11.s
C 15	280.6	10.6	435	10	AI076022
C 16	280.6	10.6	464	10	AI086029 oy70g08.x
C 17	280.6	10.6	530	10	AI239858 qh33a01.x
C 18	280.6	10.6	535	10	AI439666 tc91h03.x
C 19	280.2	10.6	356	10	AI933814 wn92f07.x
C 20	280	10.6	771	10	AI990864 ws32g06.x
C 21	279.6	10.5	467	10	AI356863 qy18a06.x
C 22	279.6	10.5	535	11	BE855380 7g13a05.x
C 23	279	10.5	398	10	AI084212 oy72e10.x
C 24	279	10.5	418	10	AI657120 tt49q11.x
C 25	279	10.5	454	10	AW592092 hf38f01.x
C 26	279	10.5	485	10	AI378090 tc80a08.x
C 27	279	10.5	477	10	AI804631 tc81q07.x
C 28	279	10.5	488	10	AI364917 qz23a06.x
C 29	279	10.5	488	10	AI802729 wf17f06.x
C 30	279	10.5	551	10	BE501539 bw33e06.x
C 31	278.4	10.5	466	11	BF438587 nab94g01.x
C 32	277.4	10.5	323	10	AI633917 tt23a09.x
C 33	277.4	10.5	490	11	BF510804 UI-H-B14-
C 34	277	10.4	292	13	AF166290 AF166290
C 35	276.8	10.4	566	11	BF064049 7h93b04.x
C 36	276.4	10.4	563	10	AI161212 qb55b11.x
C 37	276.2	10.4	508	10	AI524254 th11f03.x
C 38	275.8	10.4	328	10	AI633353 ts96b09.x
C 39	275.8	10.4	467	10	AI356259 qz25e02.x
C 40	275.8	10.4	478	10	BE674817 7e95b08.x
C 41	275.8	10.4	497	10	BE675439 7f09b05.x
C 42	275	10.4	594	10	AI914983 wf81e12.x
C 43	274.6	10.3	507	10	AI218309 qh16b07.x
C 44	274.2	10.3	461	10	BE646379 7e85h01.x
C 45	274.2	10.3	547	10	AI682378 wc52e03.x

ALIGNMENTS

RESULT 1
AQ223799/c
LOCUS HS_2218_A1_C06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2218 Col-11 Row=E, DNA sequence.
DEFINITION
ACCESSION AQ223799
VERSION AQ223799.1 GI:3639642
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 538)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 2218 row: E column: 11
Class: BAC ends
High quality sequence stop: 538.
LOCATION/Qualifiers
1..538
/organism="Homo sapiens"
/db_xref="taxon:9606"

```
/clone="Plate=2218 Col=11 Row=E"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      116 a   144 c   117 g   149 t   12 others
ORIGIN

Query Match      15.8%; Score 420.4; DB 13; Length 538;
Best Local Similarity 94.8%; Pred. No. 4.3e-46;
Matches 455; Conservative 0; Mismatches 20; Indels 5; Gaps 2;

Qy 1971 cgaagcgcgcactgactccagctggactacagagcagactctatctcaaaaaa 2030
Db 520 CCAGACGGCGCA-TGCATCCAGCTGGACTAGACGAGAGATCATCTCAAAAAA 462
Qy 2031 aaaaaaaagaaagtaacttaggtgcagggtgctctgtttattcaactgagaccgtgcc 2090
Db 461 NAANAAAAGA---GTACTTAGGTGCAGGGTGTCTCTGTATTCACTGAGACCGTGCC 406
Qy 2091 cggttatgaggtttacacagaacaaagatttcaactatgcacatttcaacgtcacc 2150
Db 405 GGGTATGAGGTTTACAGAAAGCAAGTATTCATATGCACATATTCACCGCTCACT 346
Qy 2151 tagcattgaagccagcctgtagcctgaaagcctttgtttgagggcaggtctttccca 2210
Db 345 TAGCATTGAAGCCAGCCTGTAGCCTGANAGCCTTTGCTTGAGGGCAGGTCTTTCCCAA 286
Qy 2211 aatcagacacgaagggcgaagtgaaagtgacgtccagtgcttgcaaaagatgttaactgtcac 2270
Db 285 AATGCACACAGAGAGGTGCAAGGTGAAGTGCACGTGCCATCTTGCAAAATATGTAACTGTGAC 226
Qy 2271 gaagccacagtgaggcagagagactgtccacattgcgaagtactatgtgagacg 2330
Db 225 GAAGCCCCGAGCGAGGTAGACTGTCCACATTTCCGGAAGTGCGTATGTGAGGACG 166
Qy 2331 ggggagggcgtcccttagagataagagacaataataagggagatcatcagagaaaaatcg 2390
Db 165 GGGGAGGGGCTCCCTTAGAGATAAGATACAATCAATGAAGGGAGATATCAGAGAAATCG 106
Qy 2391 taagggagagagatggtttcaagagaataagctgacacataagaagcctggcagaagcctt 2450
Db 105 TAAGGAACAGATGGTGTGTCAAGAGAATAGGCTGACCATCGAAGTACTGCGCAGGAGC 46

RESULT 2
A1760998/c 476 bp mRNA EST 20-DEC-1999
LOCUS
DEFINITION
wh97c08.x1 NCI-CGAP_CLL1 Homo sapiens cdna clone IMAGE:2388686 3',
similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mRNA
sequence.
ACCESSION
A1760998
VERSION
A1760998.1 GI:5176665
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Cloning by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
```

```
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1388 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 388.
Location/Qualifiers
1. 476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2388686"
/tissue_lib="NCI-CGAP_CLL1"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCACTCTGAAGTGGAGCGGCGGCATCTTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT      70 a   152 c   166 g   88 t
ORIGIN

Query Match      12.6%; Score 335.6; DB 10; Length 476;
Best Local Similarity 89.2%; Pred. No. 5.1e-35;
Matches 362; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 823 cagagagcgtgtggggggccagtggtggggggggggtctgactggcccccgtct 882
Db 412 CAGGTCGGTTCGTGACGACGACCGCCGCGGGCGGACCGCAGAGCCCCGGG 353
Qy 883 ctgcccgcagagagagcggcagcagcagcagcgtctaaacgcagcggcggtccccc 942
Db 352 AGCCCGCTCGAGGAGGAGCGGCGCAGCAGCGCTTTAAACCGCAGCGCGGTGCCCCCG 293
Qy 943 cggccgagcgcgcacctcatctctgaaccgctgcagcagagacacacagcgaagctg 1002
Db 292 GGGCGGAGGCGCGACCTCATCTCAACCGCTGCAGCGAGAGCACCAGGAGGAGCTGGCG 233
Qy 1003 tctgcccgtgagcgtggagcgcagctcccccagctagcgtctggcccttgcgcgcctg 1062
Db 232 TCTGCCGTCTGAGGCTGAGGCGCAGTCCCGAGCTAGGCTCTCGGCCCTTGGCCGCCGTG 173
Qy 1063 cctgtatatacgtgtctatagacctgagcgtgagcgcagcgcagcgcagcgcagccct 1122
Db 172 CCTGTATATAGTGTCTATAGAGCCTGGCGCTCTGGAGCGGCGAGGCGAGCCCGCCCT 113
Qy 1123 gtcagcgcgcgtcccccaccctcaataaattgtctgtgagtgagcggaggtctgca 1182
Db 112 GTCCAGCGCGGCTCCCGCCACCTCAATAATGTTGCTTGAGTGGACCGAGGCTCTGCA 53
Qy 1183 ggaatgcaggagggggggcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1228
Db 52 GGAATGCAGGAGGCGGGGCTCGGCCCGCCAGGTTATTTTAAAT 7

RESULT 3
A1540062/c 478 bp mRNA EST 13-APR-1999
LOCUS
DEFINITION
td08g08.x1 NCI-CGAP_CLL1 Homo sapiens cdna clone IMAGE:2075102 3',
similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mRNA
sequence.
ACCESSION
A1540062
VERSION
A1540062.1 GI:4457435
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 478)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 1586 Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 378.

```

FEATURES
  source
    1. .478
      Location/Qualifiers
        map_location: map: 376.
        organism="Homo sapiens"
        db_xref="taxon:9606"
        clone="IMAGE:2075102"
        clone_lib="NCI-CGAP_CLL1"
        tissue_type="B-cell, chronic lymphocytic leukemia"
        lab_host="DH10B"
        notes=Vector: pT7R3D-Pac (Pharmacia) with a modified
        polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
        was primed with a Not I - oligo(dT) primer [5',
        TGTTACCATCTGAAGTCGGAGCGCGCATCTCTTTTCTTTTCTTTTCTTTT
        T 3']; double-stranded cDNA was ligated to Eco RI
        adaptors (Pharmacia), digested with Not I and cloned into
        the Not I and Eco RI sites of the modified pT7T3 vector.
        Library is normalized, and was constructed by Bento
        Soares and M.Fatima Bonaldo."
      69 a 153 c 174 g 82 t
BASE COUNT
ORIGIN
Query Match 12.5% Score 332.6; DB 10; Length 478;
Best Local Similarity 93.5%; Pred. No. 1.3e-34;
Matches 347; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

Qy	851	gcgggggcgggcgctctgactcgcggcccgctctctgcgcgcagagagagagcgagacgc	910
Db	375	gccaggcgcggcgagaccgcagagcccgaggcccgctcgaggaggagcgagacgc	316
Qy	911	agcgcttaaacgcagcgcggcggtgcccccgggcgagcgagcgagcctcatcctgaacc	970
Db	315	agcgctctaaacgcagcgcggcggtgcgcgcggcgagcgcgacgtctatcctgaacc	256
Qy	971	gcctgacgcagagaccacaagcggaagctggcgctcgcgtctgaagcgtggagcgcaatcc	1030
Db	255	gctgcacgcgagagaccacaagcggaagctggcgctcgcgtctgaagcgtggagcgcaatcc	196
Qy	1031	ccagctagcgtctcgcgccttcgcgcgcgcgtgcgtgtatatacgtgtctatagagcctg	1090
Db	195	ccagctagcgtctcgcgcgccttcgcgcgcgcgtgcgtgtatatacgtgtctatagagcctg	136
Qy	1091	gcgtctgagcgagcgagcgagcccgaccctctccagcgcggtcccgccgcacccctcaat	1150
Db	135	cggtctggagcgcgagggcagcccccgaccctctgctccagcgcggtcccgccacccctcaat	76
Qy	1151	aaatggtctgtgagtgagacccgaggtctcgaggaatgcagggagggcggggtccgcgcc	1210
Db	75	aaatgttgcttgagtgagacccgaggtctcgaggaatgcagggagggcggggtccgcgcc	16

Qy	1211	cagggttattt	1221
Db	15	CAGGTTATT	5
RESULT 4			
AW515919/c		594 bp	mRNA
LOCUS		AW515919	EST
			03-MAR-2000

```

DEFINITION      xy02c11.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2851988 3'
                  similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1 (HUMAN);, mRNA
                  sequence.
ACCESSION       AW515919
VERSION         AW515919.1
KEYWORDS        GI:7154001
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 594)
AUTHORS        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL         Unpublished (1997)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgaps-r@mail.nih.gov
                Life Technologies catalog #: 11547-015
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                image.llnl.gov/image/html/iresources.shtml
                Seq primer: -400P from Glbco
                High quality sequence stop: 425.
                Location/Qualifiers
                . 1..594
FEATURES        source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2851988"
/clone.lib="NCI_CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/notes="Organ: lymph node; Vector: pCMV-Sport6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
85 a 187 c 209 g 113 t
BASE COUNT
ORIGIN

```

Query Match	12.5%;	Score 332.6;	DB 10;	Length 594;
Best Local Similarity	93.5%;	Pred. No. 1.2e-34;		
Matches 347;	Conservative 0;	Mismatches 24;	Indels 0;	Gaps 0;
Qy	851	gcggggcgggcgtctgaactcgccccgctctctgcccgcagagagagagcgagcgagcgc	910	
Db	381	gccaggccggcgggaccgcagagcccgccggagcccgctcgaggaggagcgcgagagcc	322	
Qy	911	agcgtctctaaaccgcagcggcggtgccccgcgcggcgcagcgcacgtctatcctgaacc	970	
Db	321	agcgctctaaacgcagcggcggtgccccgcgcggcgcagcgcacgtctatcctgaacc	262	
Qy	971	gctgcagcgagagacacaaagcggaagctggcgctctgcgtgtgaggtcggagcgcagtc	1030	
Db	261	gctgcagcgagagacacaaagcggaagctggcgctctgcgtgtgaggtcggagcgcagtc	202	
Qy	1031	ccagctagcgtctcgcgcccttgcgcgcccgctgcgtgtatatacgtgttctctatagagcgtg	1090	
Db	201	ccagctagcgtctcgcgcccttgcgcgcccgctgcgtgtatatacgtgttctctatagagcgtg	142	
Qy	1091	gcgtctgagcgcggagggcagccccgcagccccctgtccagcgcggtctccgcgcacccctcaat	1150	
Db	141	gcgtctgagcgcggagggcagccccgcagccccctgtccagcgcggtctccgcgcacccctcaat	82	
Qy	1151	aaatgttctctggagtggaccagagcgtctgcaggaatcgaggagggcgcgggctccgcgcc	1210	
Db	81	aaatgttctctggagtggaccagagcgtctgcaggaatcgaggagggcgcgggctccgcgcc	22	
Qy	1211	cagggttattt 1221		
Db	21	cagggttattt 11		

RESULT 4
AW515919/C
LOCUS

tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 114 a 233 c 254 g 148 t 1 others

ORIGIN

Query Match 12.5%; Score 331.2; DB 10; Length 750;
Best Local Similarity 92.6%; Pred. No. 1.7e-34;
Matches 348; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 851 gcggggcgggcgctgactgcgcggccgctctctgcccgcagagagagagcgagcgc 910
|||||
Db 377 GCCAGGCGCGGGGACCGAGAGCCCCGGGAGCCCGCTCGAGGAGGCGCGAGACG 318
|||||
QY 911 agcgtctaaacccagcggcggtgccccgcggcgagcgccgacctcatctgaacc 970
|||||
Db 317 AGCGCTTTAAACGACGCGGGGTGCCCGCGCGAGCGCGACCTCATCTGAACC 258
|||||
QY 971 gctgcagcagagagacacaaagcgaagctggcgtctgcccgtctgaggtgagcgagtc 1030
|||||
Db 257 GCTGCAGCGAGAGACCAAGCGGAAGCTGGCGCTCGCGCTGTGAGGCTGGAGCGCAGTCC 198
|||||
QY 1031 ccagctagcgtctgcgccttgcgcggccgctgctgtatatactgttctatagagcctg 1090
|||||
Db 197 CCAGCTAGCGTCTCGGCCCTTGCGCCCGCGCGCTGTATATACGTGTCTATAGAGCGCTG 138
|||||
QY 1091 gcgtctgagcgcagagggcagcccgccgacctgtccagcgcggtcccgccacacctcaat 1150
|||||
Db 137 GCGTCTGGAGCGCGAGGCGAGCGCCCGACCGCTGTCCAGCGCGGCTCCCGCACCTCAAT 78
|||||
QY 1151 aaatgttgcgtgagtgagcagaggtctctcaggaatgcaggagggcggtccgccc 1210
|||||
Db 77 AAATGTTGCTTGGAGTGACGAGGCTCTGCAGGAATGCAGGAGGCGCGGCTCCGCC 18
|||||
QY 1211 cagggttatttttaa 1226
|||||
Db 17 CAGGGTTATTCTTAAA 2
|||||

RESULT 7
AI239613/c
LOCUS AI239613 441 bp mRNA EST 13-APR-1999
DEFINITION qb37407.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1846861.3' similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1
(HUMAN);, mRNA sequence.
ACCESSION AI239613
VERSION AI239613.1 GI:3835010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1489 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 324.
Location/Qualifiers
FEATURES
1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1846861"

/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHR, and B-cell
NCI-CCAP-GCBI) were mixed, and ss circles were used in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 65 a 145 c 152 g 79 t

ORIGIN

Query Match 12.4%; Score 330.4; DB 10; Length 441;
Best Local Similarity 93.0%; Pred. No. 2.5e-34;
Matches 346; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 851 gcggggcgggcgctgactgcgcggccgctctgcccgcagagagagcgagcgagcgc 910
|||||
Db 383 GCCAGGCGCGGGGACCGCAGAGCCCCGGGAGCCCGCTCGAGGAGGCGCGAGACG 324
|||||
QY 911 agcgtctaaacccagcggcggtgccccgcggcgagcgacctcatctgaacc 970
|||||
Db 323 AGCGCTTTAAACGACGCGCGGTGCCCGCGCGCGCGCGACCTCATCTGAACC 264
|||||
QY 971 gctgcagcagagacacaaagcgaagctggcgtctgaggtggagcgagtc 1030
|||||
Db 263 GCTGCAGCGAGAGCACCAAGCGGAAGCTGCGCTGTGCGCTGTGAGGCTGGAGCGCAGTCC 204
|||||
QY 1031 ccagctagcgtctgcgccttgcgcggccgctgctgtatatactgttctatagagcctg 1090
|||||
Db 203 CCAGCTAGCGTCTCGGCCCTTGCGCCCGCGCGCTGTATACGTGTCTATAGAGCGCTG 144
|||||
QY 1091 gcgtctgagcgcagagggcagcccgacctgtccagcgcggtcccgccacacctcaat 1150
|||||
Db 143 GCGTCTGGAGCGCGAGGCGAGCGCCCGACCGCTGTCCAGCGCGGCTCCCGCACCTCAAT 84
|||||
QY 1151 aaatgttgcgtgagtgagcagcgaggtctctgcaggaatgcaggagggcggtccgccc 1210
|||||
Db 83 AAATGTTGCTTGGAGTGACCGAGCTCTGCAGGAATGCAGGAGGCGCGGCTCCGCC 24
|||||
QY 1211 cagggttattttt 1222
|||||
Db 23 CAGGGTTATTATTT 12
|||||

RESULT 8
AI682248/c
LOCUS AI682248 561 bp mRNA EST 07-MAR-2000
DEFINITION wa71904.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2301654.3' similar to gb:M55067 NEUTROPHIL CYTOSOL FACTOR 1
(HUMAN);, mRNA sequence.
ACCESSION AI682248
VERSION AI682248.1 GI:4892430
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 561)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 625)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 705 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 460.
FEATURES
source
1..625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2158166"
/clone_lib="NCI-CGAP_CELL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTCAAGTCGAGCGCGCATCTGTTTTTTTTTTTTTTTTT
T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 88 a 198 c 221 g 117 t 1 others
Query Match 12.4%; Score 329.4; DB 10; Length 625;
Best Local Similarity 93.0%; Pred. No. 3e-34;
Matches 345; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 851 gcggggcgggcgctgactgcccgcgctctctgcccagagagagcgagacgc 910
Db 382 GCCAGGCGCGCGGACCGAGAGCCCGGAGCCCGCTCGAGAGGAGCGGACGC 323
Qy 911 agcgctctaaacgcagcgcggtgcccgcgagcgagcgagcctcatctgaacc 970
Db 322 AGCGCTCTAAACCGACCGCGGTGCCCGCGGCGAGCGCGAGCTCATCTGAACC 263
Qy 971 gctcagcgagagcaccaagcggaagctgctgctgctgctgagcgagcagctcc 1030
Db 262 GCTCAGCGAGAGACCAAGCGGAAGCTGGCTCTGCGCTGAGGCTGGAGCGCAGTCC 203
Qy 1031 ccagctagcgctcggcccttgccgccccgtgctgtatatacgtctctatagagcgtg 1090
Db 202 CCAGTAGCGCTCTCGGCCCTTGGCGCCCGCTGCTGTATACGTGTTCTATAGAGCGTG 143
Qy 1091 gcgtcagcgagcgagggcagcccgaccctgtccagcggggtcccgccaccctcaat 1150
Db 142 CGGTCTGAGCGCGGAGGCGAGCCCGACCCCTGTCCAGCGGGCTCCCGCCACCTCAAT 83
Qy 1151 aaatgttgcttgagtgaccagagctctcgaggaatgacggagggcggtcccgccc 1210
Db 82 AAATGTTGCTTGAGTGGACGAGGCTCTGCAGGAATGCAGGAGGCGCGGTCCGCC 23
Qy 1211 cagggttattt 1221
Db 22 CAGGGTAATTT 12

RESULT 11
AW977599/c 337 bp mRNA EST 02-JUN-2000
LOCUS AW977599 MAGE resequences, MAGO Homo sapiens cDNA, mRNA sequence.
DEFINITION EST389708 MAGE resequences, MAGO Homo sapiens cDNA, mRNA sequence.
ACCESSION AW977599
VERSION AW977599.1 GI:8168853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 337)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt
I.-E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 381
Seq primer: Forward.
FEATURES
source
1..337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGO"
/note="Vector: pBluescriptSKm"
BASE COUNT 60 a 111 c 108 g 58 t
ORIGIN
Query Match 12.0%; Score 318.2; DB 10; Length 337;
Best Local Similarity 97.6%; Pred. No. 1.1e-32;
Matches 323; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 892 gagagagcgagcgagcgagcgtctctaaaccgcagcgcggtgcccccgccgagcgc 951
Db 337 GAGCAGGAGCGGACGAGCAGCGCTTTAAACCGCAGCGCGGTGCCCCCGCGCGAGC 278
Qy 952 gccacatcatcctgaaccgctgcagcgagagacccaaagcggaagcgtgctgcccgtc 1011
Db 277 GCCACCTCATCTTTGAACCGTTGAAGCGAGAGACCAAGCGGAAGCTGGCGTCTGCCGTC 218
Qy 1012 tgagctggagcgagctcccgagcgtctgcgcccctgcccgcctgcccgtctgctgtat 1071
Db 217 TGAGGTTGGAGCGGAGCTCCCGCAGCTAGGTTTCGGCCCTTTCGCCCGCGCTGCTGTACAT 158
Qy 1072 acgtgtctctatagagcgtggtctgagcgcgagggcgagcccccgaccctgtccagcgc 1131
Db 157 ACGTGTTCTATAGAGCTGGCGCTCTGGAGCGCGGAGGCGAGCCCGACCCCTGTCCAGCGC 98
Qy 1132 ggtctccgcccacctcaataaattgtgtgagtgagtgagcagcgctctgcaggaatgcag 1191
Db 97 GGTCTCCGCCACCCCTCAATAAATGTTGTTGAGTGGACCGAGGCTCTGCAGGAATGCAG 38
Qy 1192 ggaagggcgggcgctccgcccaggtatttt 1222
Db 37 GGAGGGCGGGGCTCCGCCCGAGGGTTATTTT 7
RESULT 12
AW207763/c 311 bp mRNA EST 02-DEC-1999
LOCUS AW207763 UT-H-Bil-aff-b-10-0-UI.sl NCI-CGAP_Sub3 Homo sapiens cDNA clone
DEFINITION UT-H-Bil-aff-b-10-0-UI.sl NCI-CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2721379 3', mRNA sequence.
ACCESSION AW207763

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 17:18:54 ; Search time 14958.6 Seconds
(without alignments)
2926.974 Million cell updates/sec

Title: US-09-820-005-3_COPY_16200_18853
Perfect score: 2654
Sequence: 1 tdtctaggccatagcttggc.....agctacttggagggtgagg 2654

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	1934.4	72.9	230552	9	AC005098	AC005098 Homo sapi
2	1930.8	72.8	131359	9	AC004883	AC004883 Homo sapi
3	1930.8	72.8	275197	9	AC004166	AC004166 Homo sapi
4	1927.6	72.6	169604	9	AC003884	AC003884 Homo sapi
5	1735.2	65.4	124526	9	AC005080	AC005080 Homo sapi
6	1115.8	42.0	17302	9	AF184614	AF184614 Homo sapi
7	1052.4	39.7	1646	9	HS47P43S03	U60972 Homo sapien
8	1039.6	39.2	184558	2	AC092405	AC092405 Papio cyn
9	1038.6	38.8	1619	9	HS47P40S03	U61240 Homo sapien
10	1027	38.7	1619	9	HS47P41S04	U61244 Homo sapien
11	1025.4	38.6	1645	9	HS47PHOX03	U57835 Homo sapien
12	532.8	20.1	149830	2	AC027219	AC027219 Homo sapi
13	517.6	19.5	149830	2	AC027219	AC027219 Homo sapi
14	284.4	10.7	57245	2	AC068263	AC068263 Homo sapi
15	279	10.5	1340	9	AF330627	AF330627 Homo sapi
16	278	10.5	1339	9	HUMNCF1A	M25665 Human neutr
17	278	10.5	1349	9	HUMNADPHO	M55067 Human 47-KD
18	277.4	10.5	1460	9	BC002816	BC002816 Homo sapi
19	229.6	8.7	175179	9	AC005237	AC005237 Homo sapi
20	228.4	8.6	170360	2	AC084773	AC084773 Homo sapi
21	226	8.5	38429	9	AC000035	AC000035 Homo sapi
22	225	8.5	239566	9	HS4J03147	AJ003147 Homo sapi
23	223.8	8.4	100000	9	AP000502	AP000502 Homo sapi
24	223.8	8.4	109646	9	HSMHCT8S22	AF019413 Homo sapi
25	223.6	8.4	158460	2	AC022826	AC022826 Homo sapi
26	222.4	8.4	47205	2	AC087652	AC087652 Homo sapi
27	221.6	8.3	164138	9	AL162386	AL162386 Human DNA
28	221.2	8.3	135405	9	AC000025	AC000025 Homo sapi
29	221.2	8.3	149308	9	AC005527	AC005527 Homo sapi
30	221.2	8.3	318488	9	AC005529	AC005529 Homo sapi
31	220.2	8.3	78533	9	HS278N12	AL033376 Human DNA
32	219.8	8.3	100182	2	AL591863	AL591863 Homo sapi
33	219.8	8.3	172757	2	AL591848	AL591848 Homo sapi
34	219.6	8.3	134210	9	AC005052	AC005052 Homo sapi
35	219.6	8.3	176728	2	AC025544	AC025544 Homo sapi
36	218	8.2	184694	2	HS536018	AL583785 Homo sapi
37	217.4	8.2	160972	2	AC021417	AC021417 Homo sapi
38	217.4	8.2	183535	2	AP002754	AP002754 Homo sapi
39	217.4	8.2	209545	2	AC087848	AC087848 Homo sapi
40	217.2	8.2	191385	2	AC025283	AC025283 Homo sapi
41	216.8	8.2	36037	9	AC003966	AC003966 Homo sapi
42	216.8	8.2	202324	2	AC068720	AC068720 Homo sapi
43	216	8.1	142494	9	AL137119	AL137119 Human DNA
44	216	8.1	150789	9	HS21P7	AL033375 Human DNA
45	216	8.1	179849	2	AC078975	AC078975 Homo sapi

ALIGNMENTS

RESULT 1
AC005098
LOCUS AC005098 230552 bp DNA PRI 07-OCT-2000
DEFINITION Homo sapiens BAC clone CTA-350L10 from 7q11.2, complete sequence.
ACCESSION AC005098
VERSION AC005098.2 GI:9211526
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 230552)
JOURNAL Toward a complete human genome sequence
MEDLINE Sulston, J.E. and Waterston, R.
AUTHORS Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL 99063792

REFERENCE 2 (bases 1 to 230552)
AUTHORS Threlde, J., Abbott, A., Graves, T., Elliott, G. and Markovic, C.
TITLE The sequence of Homo sapiens BAC clone CTA-350L10
JOURNAL Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 230552)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

4 (bases 1 to 230552)
Waterston,R.H.
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 230552)
Waterston,R.
Direct Submission
Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.

COMMENT

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_RG350L10

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/STB/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10;
actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 222330 to 222775 is a GA
rich region. The sequence is not exact but it is believed to be
the best representation of this region. The region was sized by
PCR from clone DNA at 650 bp. The region corresponds to
restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES

source

1..230552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"

/map="7q11.2"
/clone="CTA-350L10"
/clone_lib="CITB-HS-A"
441..725
/rpt_family="Alu"
783..1149
/note="similar to EST BE000775 (NID:g8261008)"
910..1312
/note="similar to EST R22057 (NID:g776838) yh25f08.s1"
1286..1716
/note="similar to EST H72453 (NID:g1044269) ys04f07.r1"
1609..1864
/rpt_family="Alu"
1984..2271
/note="similar to EST R22113 (NID:g776894) yh25f08.r1"
2337..2682
/note="similar to EST AW383168 (NID:g6887827)"
2437..2558
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
2532..2601
/rpt_family="MIR"
3054..3367
/rpt_family="Alu"
3240..3391
/note="similar to EST AI566941 (NID:g4525393) tn25b07.x1"
3588..3845
/note="similar to EST H70120 (NID:g1040326) ys01a09.r1"
3599..4017
/note="similar to EST H89808 (NID:g1080238) yu82g03.r1"
3741..3928
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
3816..3932
/note="similar to EST AW523936 (NID:g7166321)"
3819..3932
/note="similar to EST AW503416 (NID:g7139965)"
3820..3932
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
3820..3932
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
3871..3932
/note="similar to EST AI700706 (NID:g4988606) we39e12.x1"
3876..3932
/note="similar to EST T31784 (NID:g613882)"
3889..4387
/note="similar to EST T79157 (NID:g697666) yd70b07.s1"
4109..4387
/note="similar to EST AI696334 (NID:g4984234) tw59e05.x1"
4502..4812
/rpt_family="Alu"
5041..5087
/rpt_family="Mariner"
5254..5330
/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
5265..5316
/note="similar to EST AW503416 (NID:g7139965)"
5265..5330
/note="similar to EST AI700706 (NID:g4988606) we39e12.x1"
5265..5330
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
5265..5330
/note="similar to EST AW523936 (NID:g7166321)"
5265..5330
/note="similar to EST T31784 (NID:g613882)"
5265..5348
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
5284..5330
/note="similar to EST T30644 (NID:g612742)"
5305..5330
/note="similar to EST AW993840 (NID:g8254025)"
5750..6041
/rpt_family="Alu"
5964..6127
/note="similar to EST AA078486 (NID:g1837960)"

repeat_region

misc_feature

misc_feature

misc_feature

repeat_region

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

misc_feature

University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send <mailto:ecqren@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Source and construction. This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

NEIGHBORING SEQUENCE INFORMATION:

REORDERING SEQUENCE INFORMATION.
The clone sequenced to the left is RP5-1186P10. Actual start of this clone is at base position 1 of RP4-771P4; actual end is at 131359 of RP4-771P4.

FEATURES

131992	CD	Location/Qualifiers	1. .l31359	
RES	source	/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/chromosome="7"		
		/map="7q11.21-q11.23"		
		/clone="RP4-771P4"		
		/clone_lib="RPCI-4"		
		1. .229		
	repeat_region	/rpt_family="Alu"		
		87. .549		
	misc_feature	/note="match to EST T90283 (NID:g718796) yd42c06.sl"		
		306. .810		
	misc_feature	/note="match to EST T82859 (NID:g711147) yd42b05.r1"		
		373. .854		
	misc_feature	/note="match to EST AA525914 (NID:g2267983) ni61e09.sl"		
		634. .945		
	repeat_region	/rpt_family="Alu"		
		1111. .1407		
	repeat_region	/rpt_family="Alu"		
		1649. .2063		
	misc_feature	/note="match to EST H66847 (NID:g1025587) yr71c03.sl"		
		2043. .2408		
	misc_feature	/note="match to EST H68467 (NID:g1027207) yr83h06.r1"		
		2276. .2945		
	repeat_region	/rpt_family="L1"		
		3052. .3361		
	repeat_region	/rpt_family="Alu"		
		3424. .3947		
	misc_feature	/note="match to EST N39030 (NID:g1162237) yv22d08.sl"		
		3445. .3819		
	misc_feature	/note="match to EST N48220 (NID:g1189386) yv22d08.r1"		
		4350. .4703		
	repeat_region	/rpt_family="MER4-group"		
		4716. .4992		
	repeat_region	/rpt_family="Alu"		
		5005. .5265		
	repeat_region	/rpt_family="Alu"		
		5268. .5501		
	repeat_region	/rpt_family="Alu"		
		5522. .5828		
	repeat_region	/rpt_family="MER4-group"		
		6359. .6671		
	repeat_region	/rpt_family="Alu"		
		6697. .7001		
	repeat_region	/rpt_family="Alu"		
		6740. .6981		
	misc_feature	/note="similar to EST A1346557 (NID:g4083763) qp46d02.x1"		
		7127. .7471		
	misc_feature	/note="match to EST A1149529 (NID:g3677998) qc70f05.x1"		
		7567. .7859		
	repeat_region	/rpt_family="Alu"		
		7884. .7937		
	repeat_region	/rpt_family="L1"		
0000	gene	Location/Qualifiers	1. .l31359	
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/chromosome="7"		
		/map="7q11.21-q11.23"		
		/clone="RP4-771P4"		
		/clone_lib="RPCI-4"		
		1. .229		
	repeat_region	/rpt_family="Alu"		
		87. .549		
	misc_feature	/note="match to EST T90283 (NID:g718796) yd42c06.sl"		
		306. .810		
	misc_feature	/note="match to EST T82859 (NID:g711147) yd42b05.r1"		
		373. .854		
	misc_feature	/note="match to EST AA525914 (NID:g2267983) ni61e09.sl"		
		634. .945		
	repeat_region	/rpt_family="Alu"		
		1111. .1407		
	repeat_region	/rpt_family="Alu"		
		1649. .2063		
	misc_feature	/note="match to EST H66847 (NID:g1025587) yr71c03.sl"		
		2043. .2408		
	misc_feature	/note="match to EST H68467 (NID:g1027207) yr83h06.r1"		
		2276. .2945		
	repeat_region	/rpt_family="L1"		
		3052. .3361		
	repeat_region	/rpt_family="Alu"		
		3424. .3947		
	misc_feature	/note="match to EST N39030 (NID:g1162237) yv22d08.sl"		
		3445. .3819		
	misc_feature	/note="match to EST N48220 (NID:g1189386) yv22d08.r1"		
		4350. .4703		
	repeat_region	/rpt_family="MER4-group"		
		4716. .4992		
	repeat_region	/rpt_family="Alu"		
		5005. .5265		
	repeat_region	/rpt_family="Alu"		
		5268. .5501		
	repeat_region	/rpt_family="Alu"		
		5522. .5828		
	repeat_region	/rpt_family="MER4-group"		
		6359. .6671		
	repeat_region	/rpt_family="Alu"		
		6697. .7001		
	repeat_region	/rpt_family="Alu"		
		6740. .6981		
	misc_feature	/note="similar to EST A1346557 (NID:g4083763) qp46d02.x1"		
		7127. .7471		
	misc_feature	/note="match to EST A1149529 (NID:g3677998) qc70f05.x1"		
		7567. .7859		
	repeat_region	/rpt_family="Alu"		
		7884. .7937		

Matches	2030;	Conservative	0;	Mismatches	569;	Indels	55;	Gaps	
QY	1	tgctcaggccatagcttgccagctgcgcggcgggggtctcagcctgcagagagagcag	60						
Db	49590	TGTC TAGGCCATAGCTTGCACTGCGCGGGCGGGGCTCTCAGCCTGCAGAGAGCAG	49649						
QY	61	gacctcaccgggggaaaggggtggacgcgctggccgcggtgtggggtcggcagggg	120						
Db	49650	GACCCCTACGGGGGAAAGGGGTGGACGGCCTGGCCGGGTGTGGGCTGGCACGGGG	49709						
QY	121	cggaaagaaagcgcgcatcccgggggtcttggaatgggagtcagtcaggggggtccccc	180						
Db	49710	CGAAAGAAAGCGCGCATGCCCGGGGCTTTGGGATGGCAGTCCAGGGGGGCTTCCCG	49769						
QY	181	gaggggggacacagacgaagggctggtagggcggtggaaacccgcccaggtctgct	240						
Db	49770	GAGAGGGGACGACACACGAAGGCTGGTAGGGGCTGGAAACCCGCCAGGCTCTGCT	49829						
QY	241	gcagggaaggggtccctgttcgtgacggggggcagccgctctctgtcccccgggggtc	300						
Db	49830	GCAGGCAAGGGTCCCTTGTCTGTCAGCGGGGACAGCGCCTCTTGTGCCCGGGTCTG	49889						
QY	301	agactaccggcccccctactgcccccaactctcggaccaggggtgccatctcagtcgc	360						
Db	49890	AGACTACCGGCCCCCTACTTGCCTCCCTCCACTTCTCGGACAGGGGTGCCCATCTG	49949						
QY	361	tggggcaggggcgcctcggtttgacgacgcccgttcccgctgggccaggtcgtcca	420						
Db	49950	TGGGGGACAGGGGCGCCTCGGGCTTTGACGACGCCCGCTCCCGCTGGGCCAGGTCCA	50009						
QY	421	tcgcgaacgcgcacagcatccaccagcgtcgcgaaagcgcctcagccagcagcctatc	480						
Db	50010	TCCGCAACGCACACATCCACCAGCGTTCGGGAAGCGCTCAGCCAGGAGCGCTATC	50069						
QY	481	gcgcgaacagctccgctttctcagcagcagcgcgcgcgcgcgcgcgcgcgcgcgcga	540						
Db	50070	GCCGCAACAGCTCCGCTTTCTGACGACGCGCGCCAGCGCGCGGGGACCGCAGAGA	50129						
QY	541	gccccggagcccgctcggtgagtcagcggagagggcaggaagggcgaagccttaggg	600						
Db	50130	GCCCCGGAGCCCGCTCGGTGAGTCGACGGGAGAGGGCAGGAAGGCCGCTAGAGG	50189						
QY	601	cggagtcagcgggagaggcggggccagagagcggggccagagtagcggggcgaccagag	660						
Db	50190	CGGAGTCAGCGGAGAGCGGGGCCACAGAGTATGGGCCAGAGTAGCGGGCGGACACAG	50249						
QY	661	ggcgaatcagagggagagcgggggagctgagcggggggcagagagagcagcgtcag	720						
Db	50250	GGCGGAATCAGAGGGAGAGCGGGGACTGAGGCGGGGTCTCAGAGAGAGAGCAGCGCTAG	50309						
QY	721	ggggcgagcgaacctaaagagcggagtcagagggagggcacaagcgggagcggagc	780						
Db	50310	GGGGCGAGCGCATCCCTTAGAGCGCGAGTACAGGGGAGAGGACCAAGCGGGAGGCG	50369						
QY	781	cagagcgcggagcaggagttgagaccgcggcgaggccagagcagagcgtgtggcg	840						
Db	50370	CAGAGCGCGAGCAGCAGTGTGAGAACGCGCGGGCGAGGCCAGAGCGCTGTGGCG	50429						
QY	841	gggcccagtgcgggggcgggggcgctcgtacgcgcgcgcgcgcgcgcgcgcgcgcgc	900						
Db	50430	GGGCCAGTGGCGGGGCGGGGCGCTGTGACTCGCCCCCGCTCTCTCTCTCTCTCTCTCT	50489						
QY	901	cgcgacagcgcgcgtctcaaacgcgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	960						
Db	50490	CGGACACCGAGCGCTCTTAACCGGACCGCGGGGCTCCCGCGCGCGCGCGCGCGCTC	50549						
QY	961	atcctgaacgcgtgcagcagagacccaagcgggaagcgtgcgtcgtcgtcgtcgtcgtcgt	1020						
Db	50550	ATCCTGAACCGCTGCAGCGAGAGCACCAGCGAAGCTGGCGCTCTGCGCTCTGAGGCTGG	50609						
QY	1021	agcgcagtcctccagctagcgtctcggcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1080						
Db	50610	AGCGCATGCTCCAGCTAGGCTCTCGGCCCTTCGCGCGCGCGCGCGCGCGCGCGCTCT	50669						

QY 1081 atagagcctggcgctctggagcgcgagggcagcccccagaccctgtccagcgcggtcccg 1140
|||||
Db 50670 ATAGAGCCTGGCGTCTGACCGCGAGGCGAGCCCGACCCCTGTCCAGCGCGCTCCCGC 50729
QY 1141 caccctcaataattgtctgtgagtgagcagagagctctgcaggaatgcagggagggcg 1200
|||||
Db 50730 CACCCTCAATAAATGTGCTTGGAGTGGACCGAGGCTCTGCAGGAATGCAGGAGGGCGG 50789
QY 1201 ggcctcccccaggttatcttcttaagttgagcagagggaggttctgagttctgtnnnnn 1260
|||||
Db 50790 GGCTCCGCCCCAGGGTTA-TTCTAAGTTGAGGACA-GGAGGTTGTGATTCGTG----- 50841
QY 1261 nnn 1320
Db 50842 -----CTGGGGGAAGTTC 50856
QY 1321 nnn 1380
Db 50857 AAGAGCCGAGGTCTGGTTGCATGTTGCCCTGGTCTTGGCCAAAGACAGGTTTGCACAAG 50916
QY 1381 nnn 1440
Db 50917 CCAAGTTCAGAGGAACCTCCCGGTTTCTCTGTGACCGTGTGTCAGAAACCACTGCTTG 50976
QY 1441 nnn 1500
Db 50977 GACTCTGGCGGAAGAGTCTGAAGATGGTGCACACAGTGCAGCAGGGCAGCCCTGTCTC 51036
QY 1501 nnn 1560
Db 51037 ATGACAGGAGACAGGCTGCCGTCCAGGGTGTAGGAGTGACCTCATAGCTGGGATAAAAA 51096
QY 1561 nnn 1620
Db 51097 TATATTAACTTAGTTCGGGCGGGTGGCTCAGCGCTGTAAACCCAGACTTTGGGAGA 51156
QY 1621 nnn 1680
Db 51157 CGAGGTGGGAGATCCGTTGAGCTCAGAGTTCGAGCAGCAGCCCTGGCCAACTGCTGAA 51216
QY 1681 nnn 1740
Db 51217 ACCCCATCTTACCAAAAAATATAAAAAATTAGCTGGCGCTGTGGCATGCATCAATAATCC 51276
QY 1741 nnn 1800
Db 51277 CAGCTACTGGGAGGCTGAGGCATGAGAAATCGCTTGAACCGGGGAGGAGATGTTGCACT 51336
QY 1801 nnn 1860
Db 51337 GAGCCGAGACGCGCCACTGGACTCCAGCTGGACAACATGCTGAACCCCATCTGTACC 51396
QY 1861 nnnnnnnntaaataatagctggcgctggctgagcagcagcagcagcagcagcagcagcag 1920
Db 51397 AAAAAATATAAAAAATTAGCTGGCGTGGTGGCATGCATCCAAATCCAGCTACTGGGAG 51456
QY 1921 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1980
Db 51457 GCTGAGGCATGAGATCGCTTGAACCGGGGAGGAGATGTTGCACTGAGCGGAGACGGCG 51516
QY 1981 ccactgcactccagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2040
Db 51517 CCAGTGCACCTCGAGCTGGACTACAGAGGAGACTCTATCTCAAAAAAATAAAAAA 51576
QY 2041 aaagtaactaggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2100
Db 51577 AAAGTAACTAGGTGAGCGGTGCTCTGTTATTACTGAGACCGTGCCTGGGTTATGAG 51636
QY 2101 gttgtaccagaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2160
Db 51637 GTTGTAACGAAAGCAAGTATTCATATGACACTATTTCACCGCTCACCCCTAGCATGAA 51696

QY 2161 gccagcctgtagcctgaaagccttctgttggcggcagctcttccccaaaaatgcagaca 2220
|||||
Db 51697 GCCAGCCTGTAGCCTGAAGCCTTTGCTTTGAGGCGCAGGTCTTTCCCAAAATGAGACA 51756
QY 2221 ggaaggtcaaaagtgaagctgccagctcttgcataaagatgttaactgtcacgaagccag 2280
|||||
Db 51757 CGAAGGTGCAAAAGTGAAGCTGCCAGTCTTTGCAAAAGATGTAACTTGTCACGAAGGCCAG 51816
QY 2281 actgcagaggaagcgtgtcccaacatttgcggaagtggctatatgtgaggaagcgggagcgg 2340
Db 51817 AGTGCAGGAGGAGCTGTCACATTTGCGGAAGTGGCTATGTGAGGACGGGGGAGGGG 51876
QY 2341 gtcccttagagataagagacaatcataaagggagagatatcagagaaaatcgtaagggagc 2400
Db 51877 GTCCCTTAGAGAT--GAGACAATCATAGGGGAGATATCAGAGAAAATGTAAGGGGAGC 51934
QY 2401 agatggttgcataagaagaataggtgaccatcgaagcactggaagcagcagcagcagcagc 2460
Db 51935 AGATGTTGTCAAGAGAATAGGCTGACCATCGAAGAGCTGGCAGAAAGCTTTTCAGAAAACC 51994
QY 2461 actgacgcgctgggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2520
Db 51995 ACTGACGCGCTGGCAGCAGTGGCTTAGGCTGTAATCCAGCACCTTTGGGAGGCTGACGC 52054
QY 2521 agtgaatcaacttggctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2580
Db 52055 AGGTGAATCACTTGAAGTCAAGGAGTTCAGACACCGCTGGCCCAACATGTTGAACCCCAT 52114
QY 2581 ctctacagaaataataaaattagcagcagcagcagcagcagcagcagcagcagcagcagcag 2640
Db 52115 CTCTACAGAAAATATAAAAAATTAGCAGCGCTGGTGCCACAAAGCCTAGAAATCCAGCTAC 52174
QY 2641 ttggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2654
Db 52175 TTGGGAGGCTGAGG 52188

RESULT 4
AC083884
LOCUS AC083884 169604 bp DNA PRI 23-MAY-2001
DEFINITION Homo sapiens clone RP11-813J7, complete sequence.
ACCESSION AC083884
VERSION AC083884.6 GI:14190780
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 23, 2001 this sequence version replaced gi:13431264.
Center project name: H_NH0813J07.
FEATURES
Location/Qualifiers
source
1. 169604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-813J7"
BASE COUNT 44455 a 38383 c 38794 g 47962 t
ORIGIN


```
QY 2101 gttgtaccagaagaagattattcaactatgacacattattcaccgctcaccctagcattgaa 2160
Db 112814 GTTGTACCAAGAAGCAAGTATTCACTATGACACATATTACCGCTCACCTAGCATTTGAA 112873
QY 2161 gccagcctgtagcctgaagcccttgcttggaggcaggctcttccccaaaatgcagaca 2220
Db 112874 GCCAGCCTGTAGCCTGAAGACCTTTGCTTTGAGGCGAGGCTTTTCCCCAAAATGCAGACA 112933
QY 2221 cgaagtgcaaaagtgaagctgcagctcttgcaaaagatgtaacttgccacgaagccacg 2280
Db 112934 CGAAGGTGCAAGTGAAGCTGCCAGTCTTGCAAAAGATGTACTTGTACGAAGGCCACG 112993
QY 2281 agtgcagggagagctgtgccacatttgcggaagtgcctatgtgaggaaggcgaggcg 2340
Db 112994 AGTGGCAGGAGAGCTGTCCACATTTGCGGAAGTGGCTATGTGAGGACGGGGAGCGG 113053
QY 2341 gtcccttagagataagacaatacgaagggagagatatcagagaaaatgttaaggggagc 2400
Db 113054 GTCCCTTAGAGAT--GAGACAATCATAGGGGAGATATCAGAGAAAATCGTAAGGGGAGC 113111
QY 2401 agatggtgtcagaagaaatagctgaccatcgaggactggcagagcttttcagaaaaacc 2460
Db 113112 AGATGTTGTCAGAGAATAGGCTCACCATCGAAGACTGGCAGAGCTTTTCAGAAAACC 113171
QY 2461 actgacggctgggcacagtgcttaggcctgtaatcccgacactttggagcgctgacgc 2520
Db 113172 ACTGGACGGCTGGGCACAGTGGCTTAGGCCTCTAATCCACGACCTTTGGAGGCTGACGC 113231
QY 2521 agtgaatacacttgaggtcagagtgctcagacaccagccttgccaaatggtgaaaccccat 2580
Db 113232 GGGTGAATACACTTGAGGTGAGAGTGTCCAGACCGCTGGCCAAACATGGTGAAGCCCAT 113291
QY 2581 ctctcacagaaataaaaaattagcagcggtggtggcacagcctagaaatcccgactac 2640
Db 113292 CTTCTACAGAAAATATAAAATTAGCCAGCGGTGGTGGCACAAGCCTAGTAATCCCAGCTAC 113351
QY 2641 ttgggaggtgagg 2654
Db 113352 TTGGGAGGCTGAGG 113365

RESULT 5
LOCUS AC005080 124526 bp DNA PRI 02-OCT-2000
DEFINITION Homo sapiens BAC clone CTA-269P13 from 7q11.2, complete sequence.
ACCESSION AC005080
VERSION AC005080.2 GI:7770715
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 124526)
Sulston,J.E. and Watson,R.
JOURNAL
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 124526)
AUTHORS Scott,K., Layman,D., Kalicki,J. and Harmon,G.
TITLE The sequence of Homo sapiens BAC clone CTA-269P13
JOURNAL Unpublished
AUTHORS Watson,R.H.
REFERENCE 3 (bases 1 to 124526)
AUTHORS Watson,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 124526)
Waterson,R.H.
Direct Submission
JOURNAL Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```

REFERENCE
AUTHORS
TITLE
JOURNAL

5 (bases 1 to 124526)
Waterson,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 124526)
Waterson,R.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 12, 2000 this sequence version replaced gi:3212911.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapienswatson.wustl.edu
----- Summary Statistics

Center project name: H_RG269P13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-269P13 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-754G14; the clone sequenced to the right is RP11-396K3. Actual start of this clone is at base position 1 of CTA-269P13; actual end is at base position 124526 of CTA-269P13.

The clone CTA-269P13 may contain a transposon in the growth of the clone, which is not part of the submitted sequence.

FEATURES
SOURCE

1. 124526
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-269P13"
/clone.lib="CITB-HS-A"

misc_feature

1. 138
/note="match to EST AA643257 (NID:g2568475) nr61h09.s1"

misc_feature

1. 1338
/note="match to EST D44965 (NID:g1572440)"


```
misc_feature 42. .118
/notes="similar to EST D44969 (NID:g1572444)"
STS complement(106. .396)
/db_xref="GI:3800554"
misc_feature 119. .483
/notes="similar to EST AW236834 (NID:g6569223) xm49f08.x1"
misc_feature 119. .483
/notes="similar to EST AI609282 (NID:g4618449) tw83h06.x1"
misc_feature 119. .378
/notes="similar to EST D44969 (NID:g1572444)"
misc_feature 146. .483
/notes="similar to EST N80671 (NID:g1243372) zb03a07.s1"
misc_feature 164. .483
/notes="similar to EST AW204636 (NID:g6504108)"
misc_feature 204. .206
/notes="match to EST AA643257 (NID:g2568475) nr61h09.s1"
misc_feature 306. .652
/notes="match to EST AA804480 (NID:g2875993) n233c09.s1"
misc_feature 335. .520
/notes="similar to EST AW177547 (NID:g6443584)"
misc_feature 577. .666
/notes="similar to EST AW177547 (NID:g6443584)"
repeat_region 595. .711
/rpt_family="Alu"
repeat_region 712. .740
/rpt_family="(CAAAA)n"
repeat_region 741. .898
/rpt_family="Alu"
repeat_region 899. .919
/rpt_family="AT-rich"
repeat_region 965. .1106
/rpt_family="Alu"
repeat_region 1107. .1395
/rpt_family="Alu"
repeat_region 1396. .1570
/rpt_family="Alu"
repeat_region 1571. .1662
/rpt_family="L1"
repeat_region 1665. .1817
/rpt_family="L1"
repeat_region 1834. .1856
/rpt_family="L2"
repeat_region 1888. .2189
/rpt_family="Alu"
repeat_region 2256. .2546
/rpt_family="Alu"
repeat_region 3070. .3391
/rpt_family="Alu"
repeat_region 3493. .3780
/rpt_family="Alu"
repeat_region 3792. .4101
/rpt_family="Alu"
repeat_region 4108. .4232
/rpt_family="Alu"
repeat_region 4247. .4422
/rpt_family="Alu"
repeat_region 4423. .4729
/rpt_family="Alu"
repeat_region 4730. .4866
/rpt_family="Alu"
repeat_region 4941. .5099
/rpt_family="L1"
repeat_region 5222. .5527
/rpt_family="Alu"
repeat_region 5812. .5883
/rpt_family="MER1_type"
repeat_region 5884. .6194
/rpt_family="Alu"
repeat_region 6207. .6397
/rpt_family="MaLR"
repeat_region 6398. .6469
/rpt_family="MaLR"
repeat_region 6470. .6771
```

```
repeat_region /rpt_family="Alu"
6772. .7113
/rpt_family="MaLR"
7114. .7422
/rpt_family="Alu"
7423. .7468
/rpt_family="MaLR"
7469. .7644
/rpt_family="MaLR"
7645. .7774
/rpt_family="MER1_type"
7860. .8155
/rpt_family="Alu"
8194. .8466
/rpt_family="Alu"
8479. .8599
/rpt_family="MIR"
8599. .8600
misc_feature /note="similar to EST W78139 (NID:g273876)"
8601. .8916
repeat_region /rpt_family="Alu"
8787. .8803
misc_feature /note="similar to EST AA463503 (NID:g2188387) zx72c10.s1"
8810. .9112
misc_feature

Query Match 65.4%; Score 1735.2; DB 9; Length 124526;
Best Local Similarity 74.7%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 565; Indels 54; Gaps 4;

Qy 1 tgtctaggccatagcttgccagtgccgggggctctcagcctgcagagagggcag 60
Db 122131 TGTCTAGGCCATAGCTTGGCAGTGC CGGGGGGGCTCTCAGCCTGGCAGAGAGCAG 122190

Qy 61 gacctcacggggaaggagctgacgcgcctgacgcgcgtggtggatggcaggggggctccccg 120
Db 122191 GACCTCACGGGGGAAGGGGCTGGACGCGCTGCCCGGGGTGTGGGTGTCACGGGG 122250

Qy 121 cggaggaagcggcgatgcccgggggctttgggagtgggcagtgccaggggggctccccg 180
Db 122251 CGGAAGGAAGACGGCGATGCCCGGGGGCTTTGGGGATGGGCAGTCCAGGGGGCTCCCG 122310

Qy 181 gagaggggacgacagaccgaagcgtgtagggggcgctggaagacccgccaggtctgct 240
Db 122311 GAGAGGGGACACAGACCGAAGGCTGTGTAGGGGGCTGTGAAACCGCCACAGGCTCTGCT 122370

Qy 241 gcagggaagggctcctgtgacgggggacgcgcctcttgcgcgggggctgctgc 300
Db 122371 GCAGGGCAAGGGTCTTGTGTGTGACGGGGGACGGCGCTTGTGCCCGGGGTGCTGC 122430

Qy 301 agactaccggccccctactgcccccaacttccctggaccaggggtgcccatctgagtc 360
Db 122431 AGACTACCGGGCCCCCTACTGCCCCCCCACTTCTCGGACCAAGGGGTGCCCATCTGAGTCCC 122490

Qy 361 tgggggacggggcgccccctgggctttgacgacgcgcgcctcccgctggggcaggttcgcca 420
Db 122491 TGGGGGACGGGGCGCCCTCGGGCTTTTGACGACGCGCCGCTCCCGCTGGGCAAGTCTGCTCA 122550

Qy 421 tccgcaacgcgcacagatccacacagcgtgcggaagcgcctcagcagagcgcctatc 480
Db 122551 TCCGCAACGTGCACAGCATCCACCAGCGGTCCGCGAAGCGCCTCAGCCAGGACGGCTATC 122610

Qy 481 gccgcaacagcgtccgttttctgacgacgcgcgcgcagcgcgccagcgcgcgccgacgcaga 540
Db 122611 GCCGCAACAGCGTCCGTTTTCTGACAGCAGCAGCGCGCCGCGCGGCGGACCGCAGA 122670

Qy 541 gcccgggagccgcctcgtgagtgacgaggggaggggagggaggaagggcaagccctagggg 600
Db 122671 GCCCGGGAGCCCGCTCGGTGAGTGACGGGGAGAGGGCAGGAAGGGCAAGCCCTAGGGG 122730

Qy 601 cggagtcagcgaggagcggcgggccagagggcggccagagtagcggggcgggagaccagag 560
Db 122731 CGGAGTCAGCGGAGAGCGGGCGCCAGAGGTAGGGCCAGAGTAGGGGGCGGGGACCGAGAG 122790
```

QY 661 ggcggaaatcagagggagcgggactggagcggggcagagggagagccagcgctag 720
|||||
Db 122791 GCGGAAATCAGAGGAGAGCGGGACTGGAGCGGGGCGACAGAGGAGCGCTAG 122850

QY 721 gggcgagcgatccctaaagagcgagatcagagagagagcacaagcgggagcgagcg 780
|||||
Db 122851 GGGCGGAGCGATCCCTAAGAGCGGAGTCAAGAGGAGAGGACAAAGCGGGAGGAGCG 122910

QY 781 cagagcgagcagagagtgtagaccgcggcgggcgagggccagagagcgctgtggcg 840
|||||
Db 122911 CAGAGCGGACACAGAGTTGGAGACCGCGCGGGCGGAGGCCAGAGCGCTGTGGCG 122970

QY 841 gggccagtgtcggggcggggctgactgcggcccgctctctccgcagagagagag 900
|||||
Db 122971 GGGCCAGTGTGCGGGCGGGGCTGTGACTCGGCCCGCGCTCTCTGCCGCAGAGGAG 123030

QY 901 cggagagcagcgctctaaaccgagcgcgggcggtgcggcccgcgccgagcgccacetc 960
|||||
Db 123031 CGGCAGACGACGCGCTAAACCGAGCGGGCGGTCCCGCCCGCGCGAGCGCCACCTC 123090

QY 961 atcctgaaccgctgcagcgagagcaccagcggaagctggcgctcgccgtcgagcgctg 1020
|||||
Db 123091 ATCCCTGAACCGCTGACGAGAGACCAAGCGGAAGCTGGCGTCTGCCGCTGTGAGCGCTGG 123150

QY 1021 agcgagtcgccagctagctgcggcccttgccggcccgctgcctgtatatacggtttct 1080
|||||
Db 123151 AGCGCAGTCCCGACGTAGCGTCTCGGCCCTTGCCGCCCGCTGCTATATACGTGTTCT 123210

QY 1081 atagagcctggcgtgcagccgagggcagcccgagcccgacccctgtccagcgcgctccgcg 1140
|||||
Db 123211 ATAGAGCCTGGCGTGTGACGCGGAGGCGACCCGCGCCCTGTCCACGCGGCTCCCGC 123270

QY 1141 caccctcaataaattgtcttgagtgagcgagcgctctgcaggaatgcagggagggcg 1200
|||||
Db 123271 CACCCTCAATAAATTGCTTGGAGTGCACCGAGGCTCTGCAGGAATGCAGGGAGGGCGC 123330

QY 1201 ggtccgcccaggggttatttctaagtgagagcagggaggtgtgagttctgtnnnnn 1260
|||||
Db 123331 GGCTCGGCCCGAGGGTTA-TTTTCTAAGTTGAGGACA-GGAGGTTGTGAGTTCTG- 123382

QY 1261 nnn 1320

Db 123383 -----CTGGGGGGAAGTTGC 123397

QY 1321 nnn 1380

Db 123398 AAGAGCCGAGGTCGTGTCATGTTGCCCTTGCTTGCCCAAGACAGGTTTGCCACAAG 123457

QY 1381 nnn 1440

Db 123458 CCAAGTTCAAGAGGAACCTCCCGGTTTCTCTGCTGACCGTTTGTGCAGAAACCACTGCTTG 123517

QY 1441 nnn 1500

Db 123518 GACTCTGCGGAGAGTCTCTGAAGATGGGTGCACACACTGCAGAGGCGCCCTGTCTC 123577

QY 1501 nnn 1560

Db 123578 ATGGCAGGAGACAGGCTGCCGCTCCAGGGTGTAGGAGTGACCTCATAGCTGGGATAAAAA 123637

QY 1561 nnn 1620

Db 123638 TATATTATAACTTAGTTTGGCGCGGGTGGGTCAAGCTGTAAACCCAGCGCTTTGGGAGA 123697

QY 1621 nnn 1680

Db 123698 CCGAGGTGGAGATCCGTTGAGCTCAGAGTTCGAGACCGCTGGCCACATGGTGAA 123757

QY 1681 nnn 1740

Db 123758 ACCCCATCTACCAAAAAATATAAAAAATTAGTGGGCGTGGTGGCATGCCATAATCC 123817

QY 1741 nnn 1800
;

Db 123818 CAGCTACTGGGAGGCTGAGGCATGAGAAATCGCTTCAACCGGGAGGCAGATGTTGCACT 123877

QY 1801 nnn 1860

Db 123878 GAGCCGAGACGCGCCACTTGGACTCCAGCTTGCAACATGCTGAACCCCATCTGTACC 123937

QY 1861 nnnnnnnntaaaattagctggcgctggctggatcatcaccaatcccaactactcggagag 1920
|||||
Db 123938 AAAATATATAAAATTTAGCTGGCGGTGGTGGCATTCACAATCCAGCTACTTGGGGAG 123997

QY 1921 gctgagggcatgagaatcgctgaaccgggagcgagatgttcagtgagcccgagagcgcg 1980
|||||
Db 123998 GCTGAGGCATGAGAAATCGTTGAACGGGAGCGAGATGTTGCAGTGAGCCGAGAGCGG 124057

QY 1981 ccactgcaactcagcctggactacagagcgagactctatctcaaaaaaaataaaaaa 2040
|||||
Db 124058 CCACCTGCATCCAGCGCTGGACTACAGAGCGAGACTCCATCTC-AAAAAAATAAAAAA 124116

QY 2041 aaagtaacttagtgaggggtctctctgtattcactagagaccgtgcccggttatgag 2100
|||||
Db 124117 AAAGTAACCTTAGTGCAGGGTGTCTCTGTTATTCATGAGACCGTGCCTCCGTTATGAG 124176

QY 2101 gttgtaccagaagaagtattcactatgcacactattcaccgctcaccttagcattgaa 2160
|||||
Db 124177 GTTGTACCAGAAAGCAAGTATTCACTATGCACACTATTCAACCGCTCACCTAGCATGAA 124236

QY 2161 gccagcctgtagctgaaagcccttgccttgaggcgagcttctcccaaaatgcagaca 2220
|||||
Db 124237 GCCAGCCTGTAGCCCTGAAAGCCTTTGCTTTGAGGCGAGGCTCTTCCCAAAATGAGACA 124296

QY 2221 cgaaggtgcaaaagtgaagctgcagctctctgcaaaagatgtaactgtcacgaagccacg 2280
|||||
Db 124297 CGAAGGTGCAAGTGAAGCTGCCAGTCTTGCAGAAAGATGTAACCTTGTACGAAGGCCAGG 124356

QY 2281 agtgccagggagagctgtcccaatttcggaagtggctatgtgagagcgggggagcg 2340
|||||
Db 124357 AGTGCAGGGAGAGCTGCCACATTTGCGAAGTGGCTATGTGAGGACGGGGAGCGG 124416

QY 2341 gtcccttagaataagaacaatacataaaggagagatatcagaaaaatcgtaaagggagc 2400
|||||
Db 124417 GTCCCTTAGAGATAGACAAATCATAGGGGAGATATCAGAGAAATCGTAAGGGGAGC 124476

QY 2401 agatggtgtcaagaagaatagctgaccatcgaaagactggcagaagctt 2450
|||||
Db 124477 AGATGTTGTCAAGAGATAGGCTGACCATCGAAGGACTGGCAGAAGCTT 124526

RESULT 6

AF184614 Locus AF184614 17302 bp DNA PRI 07-MAY-2001
DEFINITION Homo sapiens p47-phox (NCF1) gene, complete cds.
ACCESSION AF184614
VERSION AF184614.1 GI:6983939
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 17302)
AUTHORS Chanock, S.J., Roessler, J., Zhan, S., Hopkins, P., Lee, P., Barrett, D., Christensen, B.L., Curnutte, J.T. and Goerlach, A.
TITLE Genomic structure of the human p47-phox (NCF1) gene
JOURNAL Blood Cells Mol. Dis. 26 (1), 37-46 (2000)
MEDLINE 20238075
PUBMED 10772875
REFERENCE 2 (bases 1 to 17302)
AUTHORS Chanock, S.J., Roessler, J., Zhan, S., Hopkins, P., Lee, P., Barrett, D., Christensen, B.L., Curnutte, J.T. and Goerlach, A.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Pediatric Oncology Branch, NCI, NIH/10/13N240, Bethesda, MD 20892, USA
FEATURES Location/Qualifiers


```
source          1. .17302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23"
/clone="L24; P42"
join(<1961. .2032,5227. .5307,7041. .7116,7216. .7381,
8741. .8796,10900. .11022,11486. .11593,13142. .13259,
15946. .16050,16522. .16667,17002. .>17123)
/gene="NCF1"
/product="p47-phox"
<1961. .>17123
/gene="NCF1"
join(<1961. .2032,5227. .5307,7041. .7116,7216. .7381,
8741. .8796,10900. .11022,11486. .11593,13142. .13259,
15946. .16050,16522. .16667,17002. .17123)
/gene="NCF1"
/codon_start=1
/product="p47-phox"
/protein_id="AAF34737.1"
/db_xref="GI:6983940"
/translacion="MGDTFFIRHIALGLFEKRFVPSOHVYVFLVKWDLSEKVVYRFE
TEIYFHTLKEMPPIEAGALINPENRIIPHPKWFDDGQRAENROGLTEYESTLM
SLPKISKPHLDLFFKVRPDDLLKPTDQTKRDEYLMKRGKSTADITGPILOS
YRAITANYKTSGSMAIISDGVVEKSESGWFMQKAKRGWIPASFLEPLDSPOE
TEDPEPNVAGEPYVAIKAYTAVEGDEVSLLGEAVEVTHKLLDGWYIRKSDVYGYF
SMYLOKSDQVSOAOROIKRGAAPRRSSIRNVHSIHORSRRLSODAVRNSVFLQO
RRQARQPQSPSGPLSEERQOTORSKQPAPVPPRPSADLIILNRCSESTKRLASAV"
BASE COUNT    3960 a 4611 c 4707 g 4024 t
ORIGIN

Query Match      42.0%; Score 1115.8; DB 9; Length 17302;
Best Local Similarity 98.5%; Pred. No. 3.1e-209;
Matches 1161; Conservative 0; Mismatches 7; Indels 11; Gaps 3;

Qy 1 1gtctagccatagcttgccagtcgaggcgggcggtctctcagcctggcagagagggcag 60
Db 16110 TGTCTAGGCCATAGCTTGGCAGTGC CGGGCGGGGGCTCTCAGCCTGGCAGGAGGCAG 16169

Qy 61 gacctcaccgaggggaaagggtgacgcgcctgcccgcggtggtggggtggcagcgggg 120
Db 16170 GACCCCTCAGCGGGGAAAGGGGTGACGCGCTGCGCGCGGTGTGGGCTGGACAGGGGG 16229

Qy 121 cgaagaagaagcgcgatgccggggggtttgggagatggcgagtcaggggggtctcccg 180
Db 16230 CGGAAGGAAGCGCGATGCCCGGGGGCTTTGGGATGGGAGTCCAGGGGGGTCCCGG 16289

Qy 181 gagaggggagacagaccgagcgtggtgagggcgctggaaaaaccgccagcgtctgct 240
Db 16290 GAGAGGGGACACAGACCGAAGGCTGTTGAGGGGCGTGGAAACCGCCCGAGCTCTGCT 16349

Qy 241 gcaggggcaaggtctctgtgacgggggagcgcgcctctgttcccgccgggtctgtgc 300
Db 16350 GCAGGGCAAGGCTCTTCTCGTACGGGGGACGCGCCTCTTGTCCCGCGGGTCTGTC 16409

Qy 301 agactaccggccccctactgcccccaacttctcagaccaggggtgcccattgagtcgcc 360
Db 16410 AGACTACCGGCCCCCTACTGCCCCCACTTCTCGGACAGGGGTGCCATCTGAGTCCC 16469

Qy 361 tgggggagggcgccctcgggtttgacgacgccccctccgctcggcgtggccaggtctcca 420
Db 16470 TGGGGGACAGGGGCGCCCTGCGGCTTTGACGACGCCCCCTCCCGCTGGCCAGGTCTCA 16529

Qy 421 tccgcaacgcgcacagcatccaccagcgttcgcggaagcgtcagccagcagcagcgtatc 480
Db 16530 TCCGCAACGTGCACAGCATCCACCAGCGTGC GCGAAGCGCTCAGCCAGGACGCTATC 16589

Qy 481 gccgcaacagcgtccgttttctgacgacgacgcgcgcagcgcggcgagcgcggaccgcaga 540
Db 16590 GCCGCAACAGCGTCCGTTTCTTCTGACGACGACGCGCGCGGCGCGGCGGACGCAGA 16649

Qy 541 gccccggagccgcgtcgtgagtcagcgggagagggcaggaaggcagccctagggg 600
```

Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA

FEATURES
source

Location/Qualifiers
1..1646

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="p1 clone p43"
/chromosome="7"
/map="7q11.23"

/note="Human genomic p1 clone p43 is a p47-phox pseudogene which is defined by the presence of a GT deletion at the beginning of exon 2"
Join(U60970.1:<19..7258,U60971.1:1..3348,1..1619)
/gene="NCF1"

gene

/pseudo

exon

398..502
/gene="NCF1"

exon

/number=9
974..1119
/gene="NCF1"

exon

/number=10
1454..1619
/gene="NCF1"

BASE COUNT
ORIGIN

293 a 504 c 610 g 239 t

Query Match 39.7%; Score 1052.4; DB 9; Length 1646;
Best Local Similarity 99.3%; Pred. No. 1.5e-196;
Matches 1078; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Qy 1 tgtctagggcattgttgccatgcccggggggtctcagcctcgtggcagagggcag 60
Db 562 TGTCTAGGCCATAGTTGGCAGTGGCGGGGGGGGCTCTCAGCCTGGCAGAGGCGAG 621
Qy 61 gacctccaggggaaagggctgacgcgctgcccgcgtgtggcgtgagcagggg 120
Db 622 GACCCCTACGGGGGAAAGGGGCTGACGCGCCTGGCGGGGTGGGGCTGGCAGGGGG 681
Qy 121 cggaaaggaaagcggcgatgcccgggggtttgggagtgccagtcacaggggggtccccc 180
Db 682 CGGAAGAAAGCGCGATGCCCGGGGCTTTGGGATGGGCAGTCCAGGGGGCTCCCGG 741
Qy 181 gagaggggacacagagaccgaaggtctgtagggcggtgaaacccgccaggtctgt 240
Db 742 GAGAGGGGACACAGACACCGAAGGCTGTGTAGGGGGCTGGAACCGCCAGGGCTCTGCT 801
Qy 241 gcagggaagggtcctgtgtagcgggggagccgcctctgtcccgcgggggtcgtgc 300
Db 802 GCAGGGCAGAGGTCCTGTGTGTGAGGGGGAGCGCCCTCTTTGTCGGCGGGGTGCTGC 861
Qy 301 agactaccggccccctactgcccccaactctctcggaccaggggtgcccattcgtgctcc 360
Db 862 AGACTACCGGGCCCCCTACTGCCCCCACTCTCTCGSACCAGGGGTGCCATCTCAGTCCC 921
Qy 361 tgggggcaggggcctcctggttgacgacgccccgttcccgcgtgggcccaggtcgtcca 420
Db 922 TGGGGGCAAGGGGCGCCCTGCGGCTTTGACGACGCCCTCCCGCTGGGCGCAGGTGCTCA 981
Qy 421 tccgaacgcgcacagcatccaacagcgttcggaagcgtcagcagagcgtctatc 480
Db 982 TCCGAACGTGCACAGCATCCACACGCGGTGCGGAAGCGCTCAGCCAGGACGCTATC 1041
Qy 481 gccgaacagcgtcgttttcttcagcagcagcagcgcgcgcagcgcgcgggacccgaga 540
Db 1042 GCCGAACAGCGTCGTTTCTGACAGCAGCGCCGCCAGCGGGCGGGGACCGCAGCA 1101
Qy 541 gccccggagccgcctcgtgtagtcagcgggaggggcagaaagggaagcccttagggg 600

Db 1102 GCCCCGGGAGCCCGCTCGGTAGTGCAGC-GGAGAGGGCAGGAAGGCAAGCCCTAGGGG 1160
Qy 601 cggagtcacggggagagagcggggccagagcagggccagagtagcggggcgggaccagag 660
Db 1161 CGGAGTCACGGGAGAGCGGGGCCAGAGGTAGGGCCAGAGTAGCGGGGCGGACGACGAG 1220
Qy 661 ggcgggaatcagagggagagcgggggactgagagcgggggaggggagggagcagcg-c-ta 719
Db 1221 GCGGGAATCAGAGGAGAGCGCGGGACTGGAGCGGGGGCAGAGGAGGAGCCACGCTTA 1280
Qy 720 gggggcggagcgtatcccttaagagggcggagtcagagagggagggcacaagcggggagcgag 779
Db 1281 GGGGGCGGAGCGCATCCTTAAGAGCGGAGTCAAGAGGAGAGGACAAAGCGGGAGCGCAGG 1340
Qy 780 ccagagcgcggagcagagagttggagaccgcggcgggggcggagggccagagcgtgtgggc 839
Db 1341 CCAGAGCGCGGAGCAGAGAGTTGGAGACCGCGCGGGCGGAGGCGCAGAGAGCGCTGTGGGC 1400
Qy 840 gggggcagtgctgcggggcgggcgctgactcgggccccctctctctcccgccagagagga 899
Db 1401 GGGGCCAGTGTGGGGGGCGGGCGTCTGACTCGCCCCCGCTCTCTGCCCGCAGAGGAGA 1460
Qy 900 gcggcagcgcggcgtctctaaaccgcagccggcggtgccccgcggccgcagcgccacct 959
Db 1461 GCGGCAGACGACGCGCTCTAAACCGCAGCGCGCGGTGCCCGCGCGGCGCGCACCT 1520
Qy 960 catcctgaacctgcagcagagacccaagcggaaagctggcgctcgcgctcgtgagcgtg 1019
Db 1521 CATCTTGAACCGCTGCAGCGAGAGCACCAGCGGAAGCTGTGGCGTCTGTGGCGCTG 1580
Qy 1020 gagcgagtcctccagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1079
Db 1581 GAGCGAGTCCCAGCTAGCGTCTCGGCCCTTCCGCCCGCGTGCCTGTACATACGTGTTTC 1640
Qy 1080 tataga 1085
Db 1641 TATAGA 1646

RESULT 8
AC092405/c

LOCUS

DEFINITION

AC092405

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC092405 18458 bp DNA HTG 04-JUL-2001
Papio cynocephalus anubis clone RP41-170F23, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC092405
AC092405.1 GT:14595779
HTG; HTGS_PHASE1; HTGS_DRAFT.
olive baboon.
Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Papio.
1 (bases 1 to 184558)
Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L.,
Granite, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,
Lee-Lin, S.-Q., Legaspi, R., Lim, M., Maduro, Q. L., Maduro, V.B.,
Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Prasad, A., Shevchenko, I., Snyder, B., Stancipop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 184558)
Green, E.D.
Direct Submission
Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_mouse@hgri.nih.gov

```

----- Project Information
Center project name: ccx
Center clone name: 170F23
----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178116 bases at least Q40
Consensus quality: 178998 bases at least Q30
Consensus quality: 181141 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 184058; sum-of-contigs
Quality coverage: 10.5x in Q20 bases; agarose-fp
Quality coverage: 9.60x in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 13828: contig of 13828 bp in length
 * 13829 13928: gap of unknown length
 * 13929 13925: contig of 17329 bp in length
 * 13928 13927: gap of unknown length
 * 13158 42795: contig of 11438 bp in length
 * 13158 42795: contig of 11438 bp in length
 * 42796 42895: gap of unknown length
 * 42896 61486: contig of 18591 bp in length
 * 61487 61586: gap of unknown length
 * 61587 111399: contig of 50053 bp in length
 * 111640 111739: gap of unknown length
 * 111740 184558: contig of 72819 bp in length.

FEATURES

```
location/qualifications
1. .184558
/organism="Papio cynocephalus anubis"
/db_xref="taxon:955"
/clone="RP41-170F23"
/clone lib="RP41"
```

misc_feature

misc_feature

misc feature

misc feature

misc feature

0
1
2
3
4
5
6
7
8
9

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

IN COMPANY

Very Match st Local Simi

0/ct sauch

161 cagtccac

38562 CAAGCGC

221 aaaccg

38502 GGGCGG.

281 ttgtccc

[illegible]

Query Match	39.2%	Score	1039.6;	DB 2:	Length	184558;
Best Local Similarity	62.9%;	Pred. No.	1.3e-194;			
Matches 1570; Conservative	0;	Mismatches	629;	Indels	296;	Gaps 10;

Qy	161	cagtcacaggaggctccccgagaaggggacacagaccgaaggctggtgagggcgctgg	220
Db	138562	CAAAGCGGGGCGCGCCGCCAGGTACGCGGGGTTCCTCCGGGGCTGGCGGGTCTGAAG	138503
Qy	221	aaacccgccagctcgcgtcagggcaagggtccttgtctgatcagggggcagccgcctc	280
Db	138502	GGGGGCNACACGGTTTCGCTCTCTTAGGCAGCTTGCACTGCCGGGGGGGGCTCTC	138443
Qy	281	tgtccccqcgggatcgtcagactaccgcgccctactgcgcccaacttctcggacca	340

Query Match		38.8%;	Score 1028.6;	DB 9;	Length 1619;
Best Local Similarity		99.4%;	Pred. No. 6.8e-192;		
Matches 1053;		Conservative	0;	Mismatches	4; Indels 2; Gaps 2;
Qy	1	tgtctagccatagcttgcagtcgcgagcgccgggggtctcagcgtgcagagagcag	60		
Db	562	TGCTAGGCCATAGCTTGGCAGTCCCGGGCGGGGGCTCTCAGCCTGGCAGGAGGCGAG	621		
Qy	61	gacctcacgggggaaagggctgagacgccttgccgcggtgtggggctggccagcgggg	120		
Db	622	GACCTCACGGGGAAAGGGCTGGACGCGCTTGCCCGCGGTGTGGGCTGCGACGGGG	681		
Qy	121	cgaagaaagcgcgcagatgccccgggggttttgggga tggcagtcctcaggggggtccccg	180		
Db	682	CGGAAGAAAGCGCGATGCCCCGGGGCTTTTGGGGATGGCAGTCCAGGGGGGTCCCGG	741		
Qy	181	gagaggggagcagacagacgaagctggtgagggcgctggaaccgccagcgtcgt	240		
Db	742	GAGAGGGGAGCAGACAGACCGAAGGCTGGTGAAGGGGCTGGAAACCGCCAGGCTCTGCT	801		
Qy	241	gcagggaaggggtcttctgtagcgggggagcgcctcttgcgcgcgggggtcgtgc	300		
Db	802	GCAGGGCAAGGTCCTTGTCTGTAGCGGGGCGACGCGCTCTTGTCCCGCGGGTCTGTC	861		
Qy	301	agactacggccccctactgccccccacttctctgcgaccaggggtgccccatctagtcgc	360		
Db	862	AGACTACGGCCCCCTACTGCCCCCTCTCTCTCGACAGGGGTGCCATCTGAGTCCC	921		
Qy	361	tggggcaggggcgcctcggttttaacagcgccttcccgtctggccaggtcgtcca	420		
Db	922	TGGGGGAGGGGGCCCCCTCGGGCTTTGACGACGCCCTCTCCGCTGGGCGAGGTCTCCA	981		
Qy	421	tccgaacgcgcacagcatccaccagcgtctcggaagcgcctcagcaggaacccctatc	480		
Db	982	TCCGCAACGTGCACAGCATCCACCAGCGGTTCGCGAGAGCGCTCAGCCAGGACGCTATC	1041		
Qy	481	gccgaacagcgtcgtttttctgagcagcagcagcccgccgagcgccgggacccgaga	540		
Db	1042	CCCCAACAGCGTCCGTTTCTGCAGCAGGCGACGCGCCGACAGCGCGCGGACCGCAGA	1101		
Qy	541	gccccggagccgcctcggttagtgagcggagagcgagggcgaaggaagccttaggg	600		
Db	1102	GCCCCGGAGCCGCTCGGTGAGTGCAGC-GGAGAGGGCAGGAAGGCAAGCCGTAGGGG	1160		
Qy	601	cggagtcagcgggagagcggggccagggcagggccagagtagcggggcgggaccagag	660		
Db	1161	CGGAGTCAGCGGGAGAGCGGGGCCAGAGGTAGGCCAGAGTACGGGGCGGGACAGAG	1220		
Qy	661	ggcggaatcagagggagggcggggactgagcgggggagagggagagagcagcgc-ta	719		
Db	1221	GGCGGAATCAGAGGGAGAGCGGGGACTTGGAGCGGGGCGGACAGAGGAGCGCGCTTA	1280		
Qy	720	ggggcgagcgcgtccctaagagcggagtcagagggagagggacaagcgggagcgagg	779		
Db	1281	GGGGCGGAGCGATCCCTTAAGAGCGGAGTCAGAGGGAGAGGGCAACAGCGGGAGCGAGG	1340		
Qy	780	ccagagcgcgcagcagagtgtagagcgcgcggggcgagggcagagagcgtgtgggc	839		
Db	1341	CCAGAGCGCGAGCAGAGGTGGAGACCGCGCGGGCGGAGCGCAGAGAGCGCTGTGGGC	1400		
Qy	840	ggggccagtgctgcggggcgggcgctgactcggccccctctctgccccagagagga	899		
Db	1401	GGGGCAGTGTGCGGGCGGGCGCTCTGACTCGCGCCCGCTCTCTGCCCCGAGAGGAGA	1460		
Qy	900	gcggcagacgagcgccttaaacgcgagcgcgggtgcccccccgggcgagcgcgcacct	959		
Db	1461	CGGCAGACGACGCGCTCTAAACCGCAGCCGCGGCTGCCCGCGCGCGAGCGCGACCT	1520		
Qy	960	catctgaaccgcctgcagcggagagcacaacagcgaagctgcgtctgcgtctgagcgtg	1019		
Db	1521	CATCTGAACCGCTGCAGCGAGACGACCAACGGAAGCTGCGGCTCTGCCGTCTGAGGCTG	1580		
Qy	1020	gagcgcagtcceccagctagcgtctcggcccttgcgcgcc	1058		

Db	1581	GAGCGCAGTCCCCAGCTAGCGTCTCGGCCCTTGCGGCC 1619				
RESULT	10					
HS47P41S04						
LOCUS	HS47P41S04	1619 bp	DNA	PRI	07-JAN-1998	
DEFINITION	Homo sapiens p47-phox pseudogene, clone p41, exons 9-11.					
ACCESSION	U61244					
VERSION	U61244.1	GI:2754728				
KEYWORDS	.					
SEGMENT	4 of 4					
SOURCE	human.					
ORGANISM	Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1	(bases 1 to 1619)				
AUTHORS	Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B., Green,E.D., Chanock,S.J. and Curnutte,J.T.					
TITLE	A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease					
JOURNAL	J. Clin. Invest. 100 (8), 1907-1918 (1997)					
MEDLINE	97474758					
REFERENCE	2	(bases 1 to 1619)				
AUTHORS	Chanock,S.J., Roesler,J., Hopkins,J., Lee,P.L., Bassett,D.T., Christensen,B., Curnutte,J.T. and Gorlach,A.					
TITLE	Characterization of the genomic structure of the p47-phox gene					
JOURNAL	Unpublished					
REFERENCE	3	(bases 1 to 1619)				
AUTHORS	Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and Curnutte,J.T.					
TITLE	Direct Submission					
JOURNAL	Submitted (19-JUN-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA					
FEATURES	Location/Qualifiers					
source	1..1619					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="p41"					
	/chromosome="7"					
	/map="7q11.23"					
	/note="Human genomic P1 clone p41 is a p47-phox pseudogene which is defined by the presence of a GT deletion at the beginning of exon 2."					
	398..502					
exon	/pseudo					
	/number=9					
exon	974..1119					
	/pseudo					
	/number=10					
exon	1454..1619					
	/pseudo					
	/number=11					
BASE COUNT	287 a	498 c	604 g	230 t		
ORIGIN						
Query Match	38.7%;	Score 1027;	DB 9;	Length 1619;		
Best Local Similarity	99.3%;	Pred. No. 1.4e-101;				
Matches 1052;	Conservative	0;	Mismatches 5;	Indels 2;	Gaps 2;	
Qy	1	tgtctagccatagcttgcagtcgcgagcgccgggggtctcagcgtgcagagagcag	60			
Db	562	TGCTAGGCCATAGTTTGGCAGTCCCGGGCGGGGGCTCTCAGCCTGGCAGGAGGCGAG	621			
Qy	61	gacctcacgggggaaagggctgagacgccttgccgcggtgtggggctggccagtcggcggg	120			
Db	622	GACCTCACGGGGAAAGGGCTGGACGCGCTTGCCCGCGGTGTGGGCTGCGACGGGG	681			
Qy	121	cgaagaaagcgcgcagatgccccgggggttttggggatggcagtcctcaggggggtccccg	180			
Db	682	CGGAAGAAAGCGCGATGCCCCGGGGCTTTTGGGGATGGCAGTCCAGGGGGGTCCCGG	741			

Db 648 GACCCTCACGGGGAAAGGGCTGGACGCGCTGGCCGCGTGTGGGGCTGGCACGGGG 707
Qy 121 cggaggaagcggcgatgccccggggcctttgggagatgggagtcacaggggggctccccg 180
Db 708 CGGAAGAAACGGCGATGCCCGGGGCTTTGGGATGGGAGTCCACAGGGGGTCCCGG 767
Qy 181 gagagggggacacagaccgagctggtgagggcgctggaaaaaccocccaggtctgct 240
Db 768 CAGAGGGGACACAGACCGAAGCTGTGTAGGGCGGTGGAACACCGCCAGCGCTGCT 827
Qy 241 gcagggaaggttcctgtgtgacgggggagcgcgcctcttgcgcgcggggtgtgc 300
Db 828 CGAGGGAAGGGTCTTGTCTGTGAGGGGCGAGCGCGCTCTTGTCCCGCGGGTCTGTGC 887
Qy 301 agactaccggccccctactgccccccacttctcgagaccaggggtgccatctgagtc 360
Db 888 AGACTACGGGCCCCCTACTGCCCCCCTTCTCTCGGACGAGGGTGGCCATCTAGTCCC 947
Qy 361 tgggggaggggccccctcgggctttgacgacgcccccgctccgctgggccaaggtcgcca 420
Db 948 TGGGGGAGGGGCGCCCTCGCGCTTTGACGACGCGCCCTCCGCTGGGCCAGGTCTGCA 1007
Qy 421 tccgcaacgcgcacagatccacagcggtcgcggaagcgctcagccaggaacccctatc 480
Db 1008 TCCGCAACGTCACAGCATCCACAGCGGTCTCGGAAAGCGCTCAGCCAGGACGCTATC 1067
Qy 481 gcccaacagcgctcggtttctgacgacgacgcgcgcgcgcgcgcgcgcgcgcgcga 540
Db 1068 CCGGCAACAGGCTCGTTCGTTCTGACAGACGACGCGCGGCGCGGCGGACCGCAGA 1127
Qy 541 gcccgggagcccgctcggtgagtgacgagggagagggcaggaaggccaagccctaggg 600
Db 1128 GCCCGGAGCGCGCTCGTCACTGAGTGCAGC-GGAGAGGCGAGGAGGCGCTAGAGG 1186
Qy 601 cggagtcagcgaggagcggggagcagagcagggccagagtagcgggcggggaccagag 660
Db 1187 CGGAGTCAGCGGGAGAGCGGGGCGAGAGTAGGGCCAGAGTAGCGGGCGGACCGAG 1246
Qy 661 ggcggaatcagaggagagcggggactgagcggggagagagagagcagcgc-ta 719
Db 1247 GCGGATCAGAGGAGAGCGGGGACTGGAGGCGGGGTACAGAGGAGGAGCGCGCTTA 1306
Qy 720 gggggcgagcgatccctaaagcggagtcagagggagagggacagcgagggcgag 779
Db 1307 GGGGCGAGCGATCCCTAAGAGCGGAGTCAGAGGAGAGGACACAGCGGAGCGAGG 1366
Qy 780 ccagagcgcgagcagaggttgagagaccgcgcgcggcgagggcagagagcgctgtgggc 839
Db 1367 CCAGAGCGCGAGGAGGAGTGGAGACCGCGCGGGGCGAGGCCAGAGAGCGCTGTGGC 1426
Qy 840 ggggccaagtgtcgggcgggcgctctgactcgggcccgctctctgcccagagagga 899
Db 1427 GGGGCCAGTGTGGGGCGGGCGGTCTGACTCGGCCCGCGCTCTGTGCCCGCAGAGGAG 1486
Qy 900 gcggcagacgcagcgcttaaacgcagcgcggtgcccccgcgggcgagcgccgacct 959
Db 1487 CCGGACAGCGCAGCGCTCTAAACCGCAGCGCGGTGCCCGCGGGCGAGCGCGACCT 1546
Qy 960 cactctgaaccgctgcagcgagagacaaagcggaagctggtgctgctgaggtg 1019
Db 1547 CATCCTGAACCGTGCAGCGAGAGACCAAGCGGAAGTGGCGTCTGCGCTGAGGCTG 1606
Qy 1020 gagcaggtccccaagctagctctgcgccttgcgcgcc 1058
Db 1607 GAGCGAGTCCCAAGCTAGCGTCTCGGCGCCCTTGC CGCC 1645

RESULT 12
AC027219 AC027219 149830 bp DNA HTG 03-MAR-2001
LOCUS Homo sapiens clone RP11-729P19, *** SEQUENCING IN PROGRESS ***, 79
DEFINITION unordered pieces.

ACCESSION AC027219 GI:13184220
VERSION AC027219.3
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149830)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-729P19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149830)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:11610941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8346
Center clone name: 729_P_19

* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1569: contig of 1569 bp in length
* 1570 1669: gap of 100 bp
* 1670 2364: contig of 695 bp in length
* 2365 2464: gap of 100 bp
* 2465 3099: contig of 635 bp in length
* 3100 3199: gap of 100 bp
* 3200 3762: contig of 563 bp in length
* 3763 3862: gap of 100 bp
* 3863 4534: contig of 672 bp in length
* 4535 4634: gap of 100 bp
* 4635 5474: contig of 840 bp in length
* 5475 5574: gap of 100 bp
* 5575 6460: contig of 886 bp in length
* 6461 6560: gap of 100 bp
* 6561 7785: contig of 1225 bp in length
* 7786 7885: gap of 100 bp


```
|||||
Db 38269 GCAGGCAAGGTCCTTGTCTGACGGGGGAGCGGCTCTTGTCTCCCGGGGTCTGTGC 38328
Qy 301 agactacgccccctactgccccccacttctcgcagccaggggtgccccatctgagtcgcc 360
|||||
Db 38329 AGACTACGGCCCCCTACTGCCGCCCTTCTCTGGACCGGGGTGCCATCTGAGTCCC 38388
Qy 361 tgggggaggggccccctcgggttttgcagacgccccgcgtccgctggtggccagtcgtccca 420
|||||
Db 38389 TGGGGGAGGGGCGCCCTCGGGCTTTGACGACGCGCGCTCGCGTGGCGAGGTCGTCCA 38448
Qy 421 tcgcacagcgcacagcatccaccagcgggtcgcgagagcagcctcagcagacacattatc 480
|||||
Db 38449 TCCGCAACGGCCACAGCATCATCAGCGGTCTCGCGAGAGCGCCTCAGCCAGACGCCCTATC 38508
Qy 481 gcccaacagcgtcgtcttctcgcagcagcagcgcgccccagcggcgccgggacccg 536
|||||
Db 38509 GCCCAACAGGTCGCTTCTGTCGACGAGCGACGCCGCGGCGCGCGGACCG 38564

RESULT 13
AC027219/c
LOCUS AC027219 149830 bp DNA HTG 03-MAR-2001
DEFINITION Homo sapiens clone RP11-729P19, *** SEQUENCING IN PROGRESS ***, 79
unordered pieces.
ACCESSION AC027219
VERSION AC027219.3 GI:13184220
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149830)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beka,F.,
Boguslavskiy,L., Boukhvalter,B., Brown,A., Buckett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Gadigan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lenocky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:11610941.
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8346
Center clone name: 729_P_19
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1569: contig of 1569 bp in length
* 1570 1669: gap of 100 bp
* 1670 2364: contig of 695 bp in length
* 2365 2464: gap of 100 bp
* 2465 3099: contig of 635 bp in length
* 3100 3199: gap of 100 bp
* 3200 3762: contig of 563 bp in length
* 3763 3862: gap of 100 bp
* 3863 4534: contig of 672 bp in length
* 4535 4634: gap of 100 bp
* 4635 5474: contig of 840 bp in length
* 5475 5574: gap of 100 bp
* 5575 6460: contig of 886 bp in length
* 6461 6560: gap of 100 bp
* 6561 7785: contig of 1225 bp in length
* 7786 7885: gap of 100 bp
* 7886 8835: contig of 850 bp in length
* 8836 8864: gap of 100 bp
* 8865 9964: contig of 1029 bp in length
* 9965 10784: contig of 820 bp in length
* 10785 10884: gap of 100 bp
* 10885 11653: contig of 769 bp in length
* 11654 11753: gap of 100 bp
* 11754 12415: contig of 662 bp in length
* 12416 12515: gap of 100 bp
* 12516 13232: contig of 717 bp in length
* 13233 13332: gap of 100 bp
* 13333 14178: contig of 846 bp in length
* 14179 14278: gap of 100 bp
* 14279 15762: contig of 1484 bp in length
* 15763 15862: gap of 100 bp
* 15863 16423: contig of 561 bp in length
* 16424 16523: gap of 100 bp
* 16524 17467: contig of 944 bp in length
* 17468 17567: gap of 100 bp
* 17568 18533: contig of 966 bp in length
* 18534 18633: gap of 100 bp
* 18634 19882: contig of 1249 bp in length
* 19883 19982: gap of 100 bp
* 19983 21070: contig of 1088 bp in length
* 21071 21170: gap of 100 bp
* 21171 22153: contig of 982 bp in length
* 22153 22232: gap of 100 bp
* 22253 23303: contig of 1051 bp in length
* 23304 23403: gap of 100 bp
* 23404 24381: contig of 978 bp in length
* 24382 24481: gap of 100 bp
* 24482 26183: contig of 1702 bp in length
* 26184 26283: gap of 100 bp
* 26284 27663: contig of 1380 bp in length
* 27664 27763: gap of 100 bp
* 27764 28566: contig of 803 bp in length
* 28567 28666: gap of 100 bp
* 28667 30309: contig of 1643 bp in length
* 30310 30409: gap of 100 bp
* 30410 31434: contig of 1025 bp in length
* 31435 31534: gap of 100 bp
* 31535 32225: contig of 691 bp in length
* 32226 32325: gap of 100 bp
* 32326 33690: contig of 1365 bp in length
* 33691 33790: gap of 100 bp
* 33791 34434: contig of 644 bp in length
* 34435 34534: gap of 100 bp

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9566

Center clone name: 225_K_6

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 664 763: gap of 100 bp
* 764 1454: contig of 691 bp in length
* 1455 1554: gap of 100 bp
* 1555 2227: contig of 673 bp in length
* 2228 2327: gap of 100 bp
* 2328 2833: contig of 506 bp in length
* 2834 2933: gap of 100 bp
* 2934 3614: contig of 681 bp in length
* 3615 3714: gap of 100 bp
* 3715 4404: contig of 690 bp in length
* 4405 4504: gap of 100 bp
* 4505 5193: contig of 689 bp in length
* 5194 5293: gap of 100 bp
* 5294 5959: contig of 666 bp in length
* 5960 6059: gap of 100 bp
* 6060 6717: contig of 658 bp in length
* 6718 6817: gap of 100 bp
* 6818 7486: contig of 669 bp in length
* 7487 7586: gap of 100 bp
* 7587 8278: contig of 692 bp in length
* 8279 8378: gap of 100 bp
* 8379 9040: contig of 662 bp in length
* 9041 9140: gap of 100 bp
* 9141 9808: contig of 668 bp in length
* 9809 9908: gap of 100 bp
* 9909 10573: contig of 665 bp in length
* 10574 10673: gap of 100 bp
* 10674 11375: contig of 702 bp in length
* 11376 11475: gap of 100 bp
* 11476 12157: contig of 682 bp in length
* 12158 12257: gap of 100 bp
* 12258 12954: contig of 697 bp in length
* 12955 13054: gap of 100 bp
* 13055 13746: contig of 692 bp in length
* 13747 13846: gap of 100 bp
* 13847 14524: contig of 678 bp in length

* 14525 14624: gap of 100 bp
* 14625 15267: contig of 643 bp in length
* 15268 15367: gap of 100 bp
* 15368 16052: contig of 685 bp in length
* 16053 16152: gap of 100 bp
* 16153 16832: contig of 680 bp in length
* 16833 16932: gap of 100 bp
* 16933 17596: contig of 664 bp in length
* 17597 17696: gap of 100 bp
* 17697 18373: contig of 677 bp in length
* 18374 18473: gap of 100 bp
* 18474 19145: contig of 672 bp in length
* 19146 19245: gap of 100 bp
* 19246 19917: contig of 872 bp in length
* 19918 20017: gap of 100 bp
* 20018 20705: contig of 688 bp in length
* 20706 20805: gap of 100 bp
* 20806 21500: contig of 695 bp in length
* 21501 21600: gap of 100 bp
* 21601 22261: contig of 661 bp in length
* 22262 22361: gap of 100 bp
* 22362 23019: contig of 658 bp in length
* 23020 23119: gap of 100 bp
* 23120 23786: contig of 667 bp in length
* 23787 23886: gap of 100 bp
* 23887 24571: contig of 685 bp in length
* 24572 24671: gap of 100 bp
* 24672 25338: contig of 667 bp in length
* 25339 25438: gap of 100 bp
* 25439 26102: contig of 664 bp in length
* 26103 26202: gap of 100 bp
* 26203 26863: contig of 661 bp in length
* 26864 26963: gap of 100 bp
* 26964 27643: contig of 680 bp in length
* 27644 27743: gap of 100 bp
* 27744 28444: contig of 701 bp in length
* 28445 28544: gap of 100 bp
* 28545 29219: contig of 675 bp in length
* 29220 29319: gap of 100 bp
* 29320 30005: contig of 686 bp in length
* 30006 30105: gap of 100 bp
* 30106 30783: contig of 684 bp in length
* 30790 30889: gap of 100 bp
* 30890 31572: contig of 683 bp in length
* 31573 31672: gap of 100 bp
* 31673 32347: contig of 675 bp in length
* 32348 32447: gap of 100 bp
* 32448 33134: contig of 687 bp in length
* 33135 33234: gap of 100 bp
* 33235 33916: contig of 682 bp in length
* 33917 34016: gap of 100 bp
* 34017 34701: contig of 685 bp in length
* 34702 34801: gap of 100 bp
* 34802 35480: contig of 679 bp in length
* 35481 35580: gap of 100 bp
* 35581 36273: contig of 693 bp in length
* 36274 36373: gap of 100 bp
* 36374 37033: contig of 660 bp in length
* 37034 37133: gap of 100 bp
* 37134 37795: contig of 662 bp in length
* 37796 37895: gap of 100 bp
* 37896 38553: contig of 658 bp in length
* 38554 38653: gap of 100 bp
* 38654 39267: contig of 614 bp in length
* 39268 39367: gap of 100 bp
* 39368 40053: contig of 686 bp in length
* 40054 40153: gap of 100 bp
* 40154 40842: contig of 689 bp in length
* 40843 40942: gap of 100 bp
* 40943 41629: contig of 687 bp in length
* 41630 41729: gap of 100 bp
* 41730 42416: contig of 687 bp in length
* 42417 42516: gap of 100 bp


```
/gene="NCF1"
/notes="polymorphism; compared to sequence deposited in
GenBank Accession Number AF330626"
/replace="g"

BASE COUNT      288 a      441 c      394 g      217 t
ORIGIN

Query Match      10.5%; Score 279; DB 9; Length 1340;
Best Local Similarity 92.2%; Pred. No. 3e-45;
Matches 294; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 851 gcggggcggggctgtgactgccccgcgtctctgccccgagaggagcgagcagc 910
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1020 GCCAGGCGCGCGGACCGCAGAGCCCGGAGCCCGGCTCGAGGAGGAGCGGACACGC 1079

QY 911 agcgctctaaaccgagccgcggtgcccccgggccgagcgccgacctcactgaacc 970
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1080 AGCGCTCTAAACCGCAGCCCGCGGTGCCCGCGCGGCGAGCGCGACCTCATCTGAACC 1139

QY 971 gctgcagcgagagcaccaggaagctgcgctctgcgtctgaggtgagcgagcagtc 1030
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1140 GCTGCACGAGAGACCAACAGCGGAGCTGGCGTCTGCCGTCTGAGGCTGAGCGCCAGTCC 1199

QY 1031 ccagctagcgtctgcgccccttgccgccctgacctgtatatacgtgttctatagagcctg 1090
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1200 CCAGCTAGCGTCTCGGCCCTTGCGGCCCTTGCGCCCTGTACATACGTGTCTATAGAGCCTG 1259

QY 1091 gcgtctgagcgccgagggcgcccgaccctctccagcgcggtcccgccaccctcaat 1150
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1260 GCGTCTGGACGCCGAGGGGACGCCCGACCCCTGTCTCCAGCGGGCTCCCGCCACCCCTCAAT 1319

QY 1151 aaatgttgccttgagtgga 1169
    ||||| ||||| ||||| |||||
Db 1320 AAATGTTGCTTGGAGTGA 1338
```

Search completed: January 11, 2002, 18:02:26
Job time: 32098 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 11, 2002, 09:07:28 ; Search time 8779.43 Seconds
(without alignments)
1691.530 Million cell updates/sec

Title: US-09-820-005-1
Perfect score: 1382
Sequence: 1 cctggaagtgcaggagca.....aaaaaaaaaaaaaaaaaaaa 1382

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : EST.*
- 1: em_estfun.*
 - 2: em_esthum.*
 - 3: em_estlin.*
 - 4: em_estom.*
 - 5: em_estpl.*
 - 6: em_estba.*
 - 7: em_estro.*
 - 8: em_estov.*
 - 9: em_hic.*
 - 10: gb_est1.*
 - 11: gb_est2.*
 - 12: gb_hic.*
 - 13: gb_gss.*
 - 14: em_gss_fun.*
 - 15: em_gss_hum.*
 - 16: em_gss_inv.*
 - 17: em_gss_pln.*
 - 18: em_gss_pro.*
 - 19: em_gss_rod.*
 - 20: em_gss_vrt.*
 - 21: em_gss_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	914.6	66.2	976	10	AL582825
C 2	890.2	64.4	1029	10	AL577485
C 3	867	62.7	938	10	AL554867
C 4	843.4	61.0	952	10	AL550056
C 5	836.2	60.5	914	10	AL574020
C 6	833.6	60.3	904	10	AL583217
C 7	824.6	59.7	938	10	AL559806
C 8	813.8	58.9	895	10	AL583422
C 9	800.6	57.9	906	10	AL561079
C 10	789.8	57.1	854	10	AL560372
C 11	787.8	57.0	889	10	AL583335
C 12	775.6	56.1	954	10	AL561678

13	723.4	52.3	851	11	BG758136
14	719.2	52.0	1070	11	BF974789
15	712.2	51.5	834	10	AL581347
16	703.2	50.9	978	10	BE561619
17	701	50.7	1042	11	BF976162
18	700.4	50.7	1108	11	BG106156
19	699.4	50.6	735	11	BG696253
20	696.4	50.4	1015	11	BG755063
21	688.4	49.8	752	11	BF976456
22	683.4	49.5	977	11	BF970668
23	682.2	49.4	751	11	BG398248
24	669.6	48.5	745	11	BG756837
25	640.2	46.3	870	10	BE561667
26	635.4	46.0	846	10	AL582483
27	634.2	45.9	707	10	AL583276
28	609.6	44.1	698	10	AW340740
29	598.6	43.3	887	11	BF129021
30	593.2	42.9	660	10	BE268407
31	591	42.8	617	10	BE397098
32	586.4	42.4	813	11	BF128894
33	583.4	42.2	656	11	BI227357
34	582.2	42.1	750	10	AW072388
35	570.6	41.3	602	11	BI224931
36	560.8	40.6	877	10	AA767386
37	553.6	40.1	608	10	BE397270
38	542.4	39.2	661	10	AA742367
39	541.4	39.2	551	10	BE501539
40	536	38.8	566	11	BF064049
41	533.8	38.6	547	10	AI436579
42	526	38.1	535	10	AI439666
43	524.8	38.0	535	11	BE855380
44	524.8	38.0	625	10	AI479089
45	520	37.6	545	10	BE513934

ALIGNMENTS

RESULT 1
AL582825/c
LOCUS AL582825 976 bp mRNA EST 16-FEB-2001
DEFINITION AL582825 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL008YK23 3
prime, mRNA sequence.
ACCESSION AL582825
VERSION AL582825.1 GI:12951193
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 976)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL008YK23"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 147 a 293 c 326 g 205 t 5 others
ORIGIN

```
Query Match 66.2%; Score 914.6; DB 10; Length 976;
Best Local Similarity 97.2%; Pred. No. 6.5e-128;
Matches 950; Conservative 5; Mismatches 9; Indels 13; Gaps 2;

Qy 335 acgctcatgagctgccaccacaaagatctccgcgtgtcccccactctctcgactcttcaag 394
Db 976 AGGCTCATGAGCCGCCACCAAGATCTCCGCTGTCCCACTCTCTTCAAG 917

Qy 395 gtgcgcccctgatgacctcaagctccccagcagacacacacacacacacacacacacac 454
Db 916 GTGCGCCCTGTATGACTCAAGCTCCCAAGTCCCAAGGACACACACACACACACACACATAC 857

Qy 455 ttgatgcccagaagtgcaagagttaccgagacacacacacacacacacacacacacacac 514
Db 856 TTGATRCAAAARATGGCAAGATACCCGACACACATCACCGGCCCATCATCTCTGCAG 797

Qy 515 acgtaccgcgcattgccaactcagagagacctcggtccgcgtcgagatgctctgtccacg 574
Db 796 ACGTACCGCGCATTTGCCGACTACGAGAAGACCTCGGGCTCCGAGATGGCTGTGTCCACG 737

Qy 575 ggggacgtggtgagagtgctgagagagagagagagagagagagagagagagagagagag 634
Db 736 GGGGACGTGTTGGAGGTGCTGGAGAAGACGAGAGAGCGGTTGTGTCTGTGATGATAA 677

Qy 635 gcaagcagagctgagatcccgagcgtctctctcgagccctcgagcgtctctcgagcagc 694
Db 676 GCAAGCGAGGCTGGATCCAGCATCTCTCTCGAGCCCTTGGACAGTCTCTGACGAGACG 617

Qy 695 gaagacctgagccaaactatgcaggtgagccatacgtcgccatcaagggccctacactgct 754
Db 616 GAAGACCTGAGCCCACTATGACAGTGAGGTCATAGTCGCGCATCAAGGCCCTACACTGCT 557

Qy 755 gtgagggggagagagtgctcctctcgagagtgagagtgagagtgagagtgagagtgagag 814
Db 556 GTGAGGGGGAGAGAGTGTCTCTGCTCGAGGGTGAAGGTGTGAGGTGATTCACAAAGCTC 497

Qy 815 ctggagcgct-----ggaagacgagctcagagcgtctctctcgagcgtctctcgagc 862
Db 496 CTGGACGGCTGTGGGTGTCATCAGGAAGACCGACGTCACAGGCTACTTTCCTCTCATGTAC 437

Qy 863 ctgcacaaagtccagggcagagacgtgtccagggcccaagccagatcaagcggggcgccg 922
Db 436 CTGCAAAAGTGGGGCAAGAGCTGTCTCCAGGCCCAACGCCAGATCAAGCG-GGGGCGCGC 378

Qy 923 cccgcaggtgctcatccgcaacgcgcacagcagcagcagcagcagcagcagcagcagcagc 982
Db 377 CCCGCAAGTGTCTATCCGCAACGCGCACAGATPCCATCAGCGGTGCGGAAGCGGCTC 318

Qy 983 agccagagcgcctatcgccgcaacagcgtctctctctctctctctctctctctctctctct 1042
Db 317 AGCCAGGAGCCTATCGCGCGCAACAGCGTCTGTTTCTGACAGAGCGACGCGCGCAGCGC 258

Qy 1043 cggccggagccagcagagccccgggagcccgctcgagggagggcagcagcagcagcagctct 1102
Db 257 CGGCGGAGCCGACGAGAGCCCGGGAGCCGCTCGAGGAGGAGCGGCGACAGCGCGCTCT 198

Qy 1103 aaacgcagcggcgggtgtccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1162
Db 197 AAACGCAGCGCGGGTGTCCCCGGGGCGGAGCGCGGACCTTCCTCTGAACCGTGTGACG 138

Qy 1163 gagagcaccagcggaaagctgtggtctgtcgcgtctgagagctgagcagcagcagcagcagc 1222
Db 137 GAGAGCACCAAGCGGAAGCTGGGCTGTGCGCTCTGAGGCTGGAGCGGAGTCCCCAGCTAG 78

Qy 1223 cgtctggcccttgccgcccccgctgctgtatatacgtgttctatagagcgtggtctctg 1282
Db 1223 cgtctggcccttgccgcccccgctgctgtatatacgtgttctatagagcgtggtctctg
```

Db 77 CGTCTGGCCCTTGGCGCCCGTGTCTGTATACATAGTGTCTTATAGAKCCTGGCGTCTGG 18
Qy 1283 acgcgcagggcagcagccc 1299
Db 17 ACGCGGAGGCGAGCCCC 1

RESULT 2
AL577485/c 1029 bp mRNA EST 16-FEB-2001
LOCUS AL577485 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0DI087YH09 3
DEFINITION prime, mRNA sequence.

ACCESSION AL577485
VERSION AL577485.1 GI:12940661
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1029)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1. 1029
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI087YH09"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 148 a 293 c 359 g 209 t 20 others
ORIGIN

Query Match 64.4%; Score 890.2; DB 10; Length 1029;
Best Local Similarity 93.4%; Pred. No. 2.8e-124;
Matches 961; Conservative 8; Mismatches 43; Indels 17; Gaps 4;

Qy 310 gggcacactaccagtgactgcagcagctcatgagcgtgcccacca-agatctccgcgt 368
Db 1029 GGGMAAACTTACCAGAACTCGGCACGCTCATGAGCTGCGCCACCCGAGTCTCCGCT 970

Qy 369 gtccccctctcgtactcttcaaggtgcgcctctgagcctcaagctcccccgaca 428
Db 969 GTCCCCACCTCTCTACCTTCTTCAGGTGCGCCCTGTGACTTCAAGCTCCCCACGGACA 910

Qy 429 accagacaaaaaacccagagacatacttgatgcccaagatggcaagtagtaccgcgacag 488
Db 909 ACCAGACAAAAAGCCAGAGACAWACTTGTATGTCCTCAAAAGATGGCAAGAGTACCGGACAA 850

Qy 489 acatcacccgcccatactctgcagacgtaccgcgcctatgcccactacagagaacct 548
Db 849 ACATMACCGGCCCATCATCTCGGAGACGTACCGCGCATTTGCCAATTCACGAGAGAGCT 790

Qy 549 cgggctccagatgctctgtccaggggagcgtggtgaggtctgtagagagagcagaga 608
Db 789 CGGCTCCGAGATGGCTCTGTCCACGGGGAGCTGGTGGAGGTCTGTAGAGAGACGGAGA 730

Qy 609 gcggtgtgtgtctgtcagatgaaagcgaagcgaaggtggtggtatccagcgtctctctcg 668


```

Db 721 GGACGAGGTGTCCTGCTCGAGGGTGAAGCTTTTCAGGTTCATTCACAGCTCCTCGACGG 780
Qy 823 ct-----ggaagacagctcacaggctacttcccgctccatgtacctgcaaaa 870
Db 781 GTGGTGGGTTCATCAGGAAGAGAGAGTACAGGCTTCTCCGCTCCATGTACCTGCAAAA 840
Qy 871 gtccagggcaagcgtgtccagggcccaagccagatcaagggggcgccgcccgcag 930
Db 841 GTCAGGGCAAGAGTGTCTCCAGGCCCAAGCCAGCAAGAGGGGGGGGGC--GCCCGCAG 898
Qy 931 gtctccatccgcagcaagcgacacatccaccagcggtcgcg 972
Db 899 GTCGTCCATCCGCAA--CGGCACASATCCACCAGCGGTGCGG 938

```

RESULT 4

```

AL550056 AL550056 952 bp mRNA EST 16-FEB-2001
LOCUS AL550056 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI053YF12 5
DEFINITION prime, mRNA sequence.
ACCESSION AL550056
VERSION AL550056.1 GI:12886644
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 952)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

```

```

FEATURES
source Location/Qualifiers
1. 952

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI053YF12"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT 223 a 299 c 263 g 163 t 4 others
ORIGIN

```

```

Query Match 61.08; Score 843.4; DB 10; Length 952;
Best Local Similarity 95.18; Pred. No. 2.9e-117;
Matches 931; Conservative 2; Mismatches 15; Indels 31; Gaps 5;

```

```

Qy 48 ccttcacgtcacatgcctctgtggctttgagagcgtctgtaccagccagcact 107
Db 1 CCTTCATCCGTCACATCCCTGCTGGGCTTTGAGAGCGGCTTCGTACCCAGCCAGCACT 60
Qy 108 atgtgtacatgtctcgtgtaaaaggcaggacctgtcgagaaaggtgtgtaccggcgct 167
Db 61 A--TGTACATGTTCTGTGTAATGGCAGGACCTGTCCGAGAAAGTGTCTACCGGCGCT 118
Qy 168 tcaccagatctacagttccataaaaccttaaaagaatttccctattgagggcggg 227
Db 119 TCACCGAGATCAGAGTTCATAAAACCTTAAAGAAATGTTCCCTATTGAGGCGAGGG 178

```

```

Qy 228 cgatcaatccagaaacagagatcatccccacacctccagctcccaagtgggttgacggc 287
Db 179 CcATCAATCCAGAGAACAGAGGATCATCCCCCACCCTCCCAGCTCCCCAAGTGGTTGACGGGC 238
Qy 288 agcgggccgcgcagaaacggccagggcacacttaccaggtacttgagcacgctcatgagcc 347
Db 239 AGCGGGCCGCCGAGAACACCAGGGCACACTTACCGAGTACTGCGGCACGCTCATGAGCC 298
Qy 348 tgcacacaagatctcccgctgtccccacacctctctcagcttctcgaagtgtgcctgatg 407
Db 299 TGCCCAACAAGATCTCCCGCTGTCCCACTCTTCAAGGTGGCGCCCTGATG 358
Qy 408 acctcaagctccccacgagacacagacacaaaagccagagacatacttgatgccccaaag 467
Db 359 ACCTCAAGCTCCCCACGGACACACAGACAAAAGCCAGACATACTTGATGCCCAAG 418
Qy 468 atggcaagtagtacgcgcagacacatcacggcccccatctcctgcagacgtaccgcgcca 527
Db 419 ATGGCAAGAGTACCGGACAGACATCACGGCCCCCATCATCTCGACAGAGTACCGCGCCA 478
Qy 528 ttgcaactacgagaagacctcgggctccgagatggctctgtccacggggagcgtggtgg 587
Db 479 TTGCCAACTACGAGAAGACCTTCGGGCTCCGAGATGGCTCTGTCCACGGGGGACGTTGTTG 538
Qy 588 aggtcgtagagaagcgcagagcggtgtgtgtctgtcagatgaacaaagcagagct 647
Db 539 AGTGTCTGGAGAGACGAGACGCGNTTGGTGTCTGTGCAGATGAACAAAGCGNGGCT 598
Qy 648 ggaatccagcgtctctcctcagccccctggacagctcctgcagacgagaaagaccctgagc 707
Db 599 GGATCCCCAGCATCTTCTCCTCGAGCCCCCTGGACAGTCTCTGACGAGACGGAACACCTGAGC 658
Qy 708 ccaactatgagggtagaccatcgtcccatcaaggccacactgctgtgagggggagc 767
Db 659 CCAACTATGTCAGGTGAGCCATACCTCGCCATCAAGGCCCTACACTGCTGTGAGGGGGAGC 718
Qy 768 aggtgtccctcgtcaggggtgaagctgttgaggtcattcacaaagctcctgcagcgctgga 827
Db 719 AGGTGTCTCTCTCGAGGGTGAAGCTGTGA-----GGG 752
Qy 828 aagcagcgtcacaggctacttcccgtccatgtacctgcaaaagtccaggcgaagcgtgt 887
Db 753 AAGACGAGGTATAGGCTACTTCCCGTCCATGTACCTGCAAAAGTCAGGCGAAGACGCTGT 812
Qy 888 cccagggcccaacgcagatcaagcggggggggcggccgcaggtcgtccatccgcgaacg 947
Db 813 CCCAGGCCCAACGCCAGATCAAGCGGGGGSG-CGCCGCGCAGGTGCGTCCATCGGCAACG 871
Qy 948 cgcacagcatccaccagcgggtcgcggaagcgcctcagcagcagcctatccgcgcaaca 1007
Db 872 TGCACAGCATCCACAGCGGTGCGGGAAGC--CCTCAGCCAGGACGCGCTATCGCCGC-AMA 929
Qy 1008 gcgtccgtttttctgcagca 1026
Db 930 GCGTCCGTTTTCTGCAGCA 948

```

```

RESULT 5
AL574020/c
LOCUS AL574020 914 bp mRNA EST 16-FEB-2001
DEFINITION AL574020 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI053YF12 3
prime, mRNA sequence.
ACCESSION AL574020
VERSION AL574020.1 GI:12933820
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 914)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)

```

COMMENT	FEATURES	source
Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers	1. 914 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DI053YF12" /clone_lib="LTI_NFL006_PL2" /tissue_type="placenta" /notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	137 a 273 c 305 g 194 t 5 others

BASE COUNT	137 A	273 C	305 G	194 T	5 Others
ORIGIN					
Query Match	60.5%	Score 836.2	DB 10	Length 914	
Best Local Similarity	95.6%	Pred. No. 3.5e-116			
Matches 899	Conservative 1	Mismatches 12	Indels 28	Gaps 3	
Qy	360	ttctccgctgtcccccacctctctcgaactcttccaagtgctgcgccttgatgacctcaagctcc	419		
Db	914	TCTCCGCTGTCCCCACCTCTTAACCTTATCTTAAAGTGCGCCTGATCACCTCAAGCTCC	855		
Qy	420	ccacggacaacccagacaaaaaagccagagacatacttgatccccaaagatggcagaagata	479		
Db	854	CCACGGACAAACACAGACAAAAGCCAGAGACATACTTGATCCCAAGATGGCAAGAGTA	795		
Qy	480	ccgcagacagacatcacgcggcccccatacctctgcagacgtacgcggccattgccaaactacg	539		
Db	794	CCGCAACAGACATCACCGGCCCATCATCTTCGAGACGCTACCGCGCCATTGCCAACTACG	735		
Qy	540	agaagacctcgggtccgagatggctctgtccacgggggacgtgggtgagctcgttagaga	599		
Db	734	AGAAAGCTCGGGTCCGAGATGGCTCTGTCCACGGGGGACGTGTGTGGAGGTCTGTGGAGA	675		
Qy	600	agacggagacggttggtggtctctgttcagatgaaagcaaacgagagctcgatccccagcgt	659		
Db	674	AGACGGAGCGGTGGTGGTCTGTCTAGATGAAGCAAGCGAGGCTGGATGCCAGCAT	615		
Qy	660	cttctctgacccctggacagctcgtacgacgagacggaagacctgagcccactatgcag	719		
Db	614	CCTTCTCGAGCCCTGNACAGTCTGTGACGAGCGGAAGACCTTGAGCCCAACTATGGAG	555		
Qy	720	gtgagccatacgttcgcctatcgaagcctacactgtctgtgagggggacagtgctccctgc	779		
Db	554	GTGAGCCATACGTCCGCATCAAGGCCTACACTGTGTGTGGAGGGGACGAGGTGTCCCTGC	495		
Qy	780	tcgagggtgaagctgttgtaggctcattcaaacgtccttggacggtcggaaagacgacgtca	839		
Db	494	TCGAGGTTGAAGCTGTTCGA-----GGAAAGACGACGCTCA	461		
Qy	840	caggctacttcccgttccatgtacctgcaaaagtcagggcaagacgtgtccacggcccaac	899		
Db	460	TAGGCTACTTCCCGTCCATGTACCTGTCAAAAAGTCAGGGCAAGACGCTGTCCACAGGCCAAC	401		
Qy	900	gccagatcaagcgggggggcgcgcgcctccaggttcctatccgcaacgcgcacagcat-c	958		
Db	400	GCCAGATCAAGC-GGGGGCGCGCCCGAGGTCTGTCCATCCGCAACGTGCACAGCATGC	342		
Qy	959	caccagcggctcgcggaagcgcctcaagccagagacgcctatcgcgcgaacagcgttcggttt	1018		
Db	341	CACCAAGCGTCCGGAAGCGCTCAGCCAGNACGGCTATCGCCGACAGCGGCTCGCGTTT	282		

Qy	1019	ctgcagcagcagccgcccgaagcgccggacgaccgagagccccggaggccgcgtcgag	1078
Db	281	CTGCAGCAGCGACGCCGCCAGCGCCGCGGGACCGCGAGAGCCCCGGAGCGCCGTCTGAG	222
Qy	1079	gagcagcgccagcagcagcgctctataaacccgagcgccggctgcccccgccgagcgccg	1138
Db	221	GAGAGCGGCGACAGCAGCGCTCTTAACCGCAGCGCGGGTTCGCCCGCGCGAGCGCCC	162
Qy	1139	gacctcatcttgaaccgcgtgcagcagcagcaccagaagcggaagcttgccgctctga	1198
Db	161	GACCTCATCTCTGAACCGCTGCAGCGAGAGCACCAAGCGGANGCTGGCATCTGCCGCTCTGA	102
Qy	1199	ggctcggagcgccagtcgcccaagctagcgtctcggcccttgccgccccgcgtctgtatatacg	1258
Db	101	GGCTGGAGCGCAGTCCCGCAGCTAGCGTCTCGCGCCTTGCCCGCCCGCTGTATATACG	42
Qy	1259	tgtctctagagcctgcgctctgaacccgagcgagcgcc	1298
Db	41	TGTTCTATAGAGCCTGGCGTCTGGACGCCGAGGCGAGCCCC	2
RESULT	6		
AL583217	AL583217	904 bp	mRNA EST 16-FEB-2001
LOCUS	AL583217	LTI_NFL010_BC2	Homo sapiens cDNA clone CS0DL012YG22 5
DEFINITION	prime,	mRNA sequence.	
ACCESSION	AL583217		
VERSION	AL583217.1	GI:12951967	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 904)		
JOURNAL	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.		
COMMENT	Full-length cdna libraries and normalization Contact: Genoscope Genoscope - Centre National de Sequencage BP 191, 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr. Location/Qualifiers 1. 904 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DL012YG22" /clone_lib="LTI_NFL010_BC2" /sex="male" /tissue_type="B cells from Burkitt lymphoma" /note="vector: pCMVSPORT 6; Site_1: NotI; First strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 5 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
FEATURES			
source			
BASE COUNT	213 a	274 c	257 g 158 t 2 others
ORIGIN			
Query Match	60.3%;	Score 833.6;	DB 10; Length 904;
Best Local Similarity	97.1%;	Pred. No. 8.6e-116;	
Matches 873:	Conservative	1; Mismatches 11;	Indels 14; Gaps 2;
Qy	40	gggggacacctcatccgtcacatcgccctgcctggcctttgagaagcgtctcgaccag	99
Db	1	GGGGGACACTTCATCCGTCACATCGCCCTGCTGGGCTTTGAGAGCGCTTCGTACCAG	60
Qy	100	cagcactatgtgtacattgttcctggtgaaatggcaggacctgtcgagagaagtgtcta	159

Db	61	CCAGCACTA--TGTACATGTTCTCTGGTAAATGGCAGGACCTGTGCGAAGAGTGGTCTA	118
Qy	160	CGGCGCTCACCAGATCTACGAGTTCATAAAACCTTAAAGAAATGTTCCCTATTGA	219
Db	119	CGGGGCTTACCGAGATCTACGAGTTCATAAAACCTTAAAGAAATGTTCCCTATTGA	178
Qy	220	ggcaggggcgatcaatccagagagacagatcatccccacacctccagctcccaagtgtt	279
Db	179	GGCAGGGCGATCAATCCAGAGAACAGGATCATCCCCACCTCCAGCTCCCAAGTGGT	238
Qy	280	tgacggcgagggcgccgagagaccgcagggcacacttaccagtagctcagcaagct	339
Db	239	TGACGGGAGCGGGCGCGGAGAAACACACAGGACATCACCAGTACTCCAGCACGCT	298
Qy	340	catgagctgcccacaaagatctcccgctgtcccccacctctcaacttctcaagtgag	399
Db	299	CATGAGCTTGGCCACCAAGATCTCCCGTGTCCCGACCTCTCGACTTCTTCAAGGTGG	358
Qy	400	ccctgatgacctcaagctcccccacggacaacacagacaaaaagccagagacatactgat	459
Db	359	CCCTGATGACCTCAAGCTCCCGACGACACACAGACAAAAAGCCAGAGACATCTTGT	418
Qy	460	gccaaagatggcaagagatccgcagacagacatcacggcccccatcatctcgcagaagta	519
Db	419	GCCCCAAGATGGCAAGAGTACCGGACAGACATCACCAGGCCCCATCATCTTGCAGAGTA	478
Qy	520	cgcgccattcccaactacgagagacctcggtctcgagatgctctgtccacgggga	579
Db	479	CGCGCCATTGCCACTACGAGAGACCTCGGGCTCCGAGATGCTCTGTCCACGGGGGA	538
Qy	580	cgtggtgaggtcgttagaagagcgagagcggttggtggtctgtcagatgaagacaaa	639
Db	539	CGTGTGAGGTCTGTGGAGAGAGCGAGAGCGNTTGTGTCTGTCTCAGATGAAGCAAA	598
Qy	640	gcgaggttgatccagcgct	699
Db	599	CGAGGCTGGATCCAGCATCTTCTCGAGCCCTGACAGTCTCTGACGAGACGAAGA	658
Qy	700	ccctgagcccaactatgcaggtgagcatatcgctcccatcaagcctcactgctgtgga	759
Db	659	CCCTGAGCCCAACTATGAGGTGAGCCATAGCTGCCATCAAGGCCCTACACTGCTGTGA	718
Qy	760	ggggagcagaggtgctccctgctcaggggtgaagctgttgagtcattcacaaagctccctgga	819
Db	719	GGGGCAGCAGGTGTCCTGCTCGAGGGTGAAGCTTTGAGGTCTTACAAAGCTCCTGGA	778
Qy	820	cggct-----ggaagacgagcgtcacaggctacttcccgctccatgtacctgca	867
Db	779	CGGCTGTGGTGTATCAGGAAAGACGAGCTACAGGCTACTTTCCGTCCTCATGTACCTGCA	838
Qy	868	aaagtcaagggcaagcgtgtcccgagggcccaacgcagatcaagcgggggcgccgccc	926
Db	839	AAAGTCGGGGCAAGAGTGTCCAGGGCCCAACGCGCARATCAAGCGGGGGGGGGCCCGC	897
RESULT	7		
AL559806			
LOCUS	AL559806	938 bp	mRNA
DEFINITION	AL559806	LTI_FL011_BC1	Homo sapiens cDNA clone CS0DG005YE05 5 prime
ACCESSION	AL559806		
VERSION	AL559806.1	GI:12905654	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		

BP 191	91006	EVRY cedex - France	
Email:	seqref@genoscope.cns.fr	Web :	www.genoscope.cns.fr.
FEATURES			
source	1.	938	
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DG005YE05"		
	/clone_lib="LTI_FL011_BC1"		
	/sex="male"		
	/tissue_type="B cells from Burkitt lymphoma"		
	/lab_host="DH10B"		
	/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	220 a	287 c	262 g
ORIGIN			5 others
Query Match	59.7%	Score 824.6;	DB 10; Length 938;
Best Local Similarity	96.6%	Pred. No. 1.9e-114;	
Matches	906;	Conservative	4; Mismatches 9; Indels 19; Gaps 6;
Qy	39	tgggggacaccttcacgtccatcgctgcctgtggtggtctttgagaagcgcttcgtaccca	98
Db	1	TGGGGGACACCTTCATCGTCCATCGCCCTGCTGGGCTTTTGAGAGCGCTTCGTACCCA	60
Qy	99	gcacgactatgtatcatgttctctgtgaaatggcaggacctctcggaagagtggtct	158
Db	61	GCCAGCACTA--TGTACATGTTCTGTGTAATGGCAGGACCTCTCGAGAAGTGTCT	118
Qy	159	accggtctcaccgagatctacagttccataaaaccttaaaagaatgttccctattg	218
Db	119	ACCGGCGCTTCACCGAGATCTACGAGTTCCTAATAAACTTTAAAGAAATGTTCCCTATTG	178
Qy	219	aggcaggggcgatcaatccagagacaggatcatccccacacctcccgagctcccaagtgt	278
Db	179	AGGCAGGGGCGATCAATCCAGAGAAACAGGATCATCCCCACCTCCAGCTCCCAAGTGGT	238
Qy	279	ttgacggcgagggcgcgagaaacgcgcagggcacacttaccagtagtactgagacgc	338
Db	239	TTGACGGGCGAGCGGGCGCGAGAACACCAGGCGACACTTACCGAGTACTGCGACGCG	298
Qy	339	tcatgagcctgccaccaaagatctcccgctgtccccacctctcgactcttccaaggtgc	398
Db	299	TCATGAGCCTGCCACCAAGATCTCCCGCTGTCCCGACCTCTCGACTTCTTCAAGGTGC	358
Qy	399	gacctgatgacctcaagctcccccacggacacaccagacacacacacacacacacacac	458
Db	359	GCCCTGATGACTCAAGCTCCCCACGGACAAACAGAC-AAAAAGCCAGAGACATCTTGA	417
Qy	459	tgcccaagagatgcaagatccgcagacacatcacggcccccatctcctgcagacgt	518
Db	418	TGCCAAAGATGGCAAGAGTACCGGACAGACATCACCGGCCCATCATCTCTTAGACGT	477
Qy	519	accggtccattgccaaactacgagagacctcggtgtcccgagatg-gctctgtccacgggg	577
Db	478	ACCGGCCATTGGCGACTACGAGAAAGACCTCGGGCTCCGAGATGTGCTCTCCACGGGG	537
Qy	578	gacgtggtgaggtcgttagaagagcgagagcggttgggtgtctctcagatgaagaca	637
Db	538	GACGTGTGGAGTCTGTGGAGAGAGCGAGCGGNTTGGTGGTCTCTCAGATGAAGACA	597
Qy	638	aagcaggtggtatccccagcgtctctctcctcgaccctcgacagtcctgcagagacgaa	697
Db	598	AAGCGAGCTGGATCCCGAGCATCTCTCTCGAGCCCTTGACAGTCTCTGACGACGGAA	657
Qy	698	gacctgagccccaaactatgcaggtgagccatacgctcgcatcaaggcctcacactgtgtg	757

```

|||||
658 GACCCCTGAGCCAACTATGAGGTGAGCCATACGTCGCCATCAAGGCTACACTGCTGTG 717
Qy 758 gagggagagaggtgctccctgctcgaaggtgaagctttaggtcattcaaacctcgt 817
Db 718 GAGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTGAGGTCAATCAAGCTCCTG 777
Qy 818 gacggct-----ggaagagagagctacaggtcacttccctgctccatgactcgt 865
Db 778 GACGGCTGGTGGTGCATCAGGAAGACGAGCTCACAGGCTACTTTCGTCATGTACCTG 837
Qy 866 caaagtcaagggaagcgtgtgtccagggcccaagccagatcaagcgggggggcccgc 925
Db 838 CAAAGTCGGGGCAAGAYGTGTCCAGGCCCCCAAGCCAGATCAAGCGGGGGGCGC--GCC 895
Qy 926 cgaaggtcgtccatccgcaacgcgcagacatccacca 963
Db 896 CGCAGGTGCTCCATCCGCAAG-SCACAGCATCCATCA 932

RESULT 8
AL583422/c 895 bp mRNA EST 16-FEB-2001
LOCUS AL583422 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL012YG22 3
DEFINITION prime, mRNA sequence.
ACCESSION AL583422
VERSION AL583422.1 GI:12952368
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 895)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
location/Qualifiers
1..895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL012YG22"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/notes="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 131 a 272 g 198 t 2 others
ORIGIN

Query Match 58.9%; Score 813.8; DB 10; Length 895;
Best Local Similarity 96.2%; Pred. No. 7.9e-113;
Matches 857; Conservative 2; Mismatches 19; Indels 13; Gaps 2;

Qy 436 aaaaagccagagacatacttgatgccaaagatggaagagtaccgcagacagatcac 495
Db 894 AAAAGCCSAGAGACATACTTGATGCCCAAGATGGCAAGAGTACCGCGCAGACATCAC 835
Qy 496 cggcccatcatctgcagacgtaccgcgcattgccaaactacagagaacctgggtc 555
Db 834 CGGCCCATCATCTCGCAGAGCTACCGCGCCATTGCCGACTACGAGAAGACCTCGGGCTC 775

```

```

556 cgaatggctctgtccacgggggagctggtgaggtcgtagagaagcagagcgggttg 615
Db 774 CGAGATGGCTCTGTGCCAGGGGACGTGGTGAGGTGCTGGGAAGAGGAGAGCGGTG 715
Qy 616 gtggttctgtatgataaagcaagcagagctggtgataccagcgtctctctcgcagccct 675
Db 714 GTGCTTCTGTGTCAGATGAAGCAAGCAGGAGGTGGATCCAGCATCTCTCTCGAGCCCT 655
Qy 676 ggaagtcctgagcagagcaagaccctgaagcccaactatcaggtgaagccatcgtcgc 735
Db 654 GGACAGTCTTACGAGACGGGAAGACCTTGACCCAACTATCGAGGTGAGCCATACGTGCG 595
Qy 736 catcaagcctacactgctgtgagggggagaggtgtccctgtctcaggttgaagctgt 795
Db 594 CATCAAGCCCTACACTGCTGTGGAGGGGACAGGTGTCCCTGCTCGAGGGTGAAGCTGT 535
Qy 796 tgaagtcattcaagaagctctctggacggct-----ggaagagacgacgtcaagg 843
Db 534 TGAGGTCAATTCACAAGCTCTCTGGACGGCTGTGGGTGTCATCAGGAAGACGACGTCA 475
Qy 844 ctacttccgtccatgtacctgcaaaagtcaggcaagacgtgtccagggcccaagcga 903
Db 474 CTACTTTCCGTCATGTACTGTCAAAAGTCGGGCAAGACGTGTCCCAAGGCCCAAGGCCA 415
Qy 904 gatcaagcgggggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 963
Db 414 GATCAAGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 355
Qy 964 gcgtgcgggaagcgtctacagcaggagcgtctatgcgcgaacagcgtccgttttctgca 1023
Db 354 GCGGTGCGGGGAAGCGCTCTACGACGAGGACGCTATCGCGCAACAGCGTCCGTTTTCTGCA 295
Qy 1024 gcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1083
Db 294 GCAGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 235
Qy 1084 gcgcagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1143
Db 234 GCGCGACGCGCGCGCGCTTAACCGCAGCGCGGTGCCCGCGCGCGCGCGCGCGCGCG 175
Qy 1144 catcctgaaccgctgcagcagagacccaagcgaagcgtggcgtcgtcgtgaggtg 1203
Db 174 CATCTTGAACCGCTGCAGCGAGAGACCAAGCGGAAGTGGCGTCTGCGCGTCTGAGGCTG 115
Qy 1204 gacgcagtcctccagctagcgtctcgccttgcgccttgcgccttgcctgtatagcttgc 1263
Db 114 GAGCGAGTCCCGAGTACGCTGTGCGCCCTTGGCGCCCGCGCGCGCGCGCGCGCGCG 55
Qy 1264 tatagacgtcgtcgtcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1313
Db 54 TATAGAGCCTAATCTTAACACCGAGRACAAACACCGACCCGCCCTATCCCAAC 4

RESULT 9
AL561079 906 bp mRNA EST 16-FEB-2001
LOCUS AL561079 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL001YL02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL561079
VERSION AL561079.1 GI:12908154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 906)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```



```
Qy 124 ggtgaaatggcaggacgtgtcggaaggtgtgtctaccggttcaccgagatctacga 183
Db 170 GGTGAAATGGCAGCACCTGTCCGAGAAGGTGTWACCGCGCTTACCAGAGATCAACGA 229
Qy 184 gttccataaaaccttaaaagaaatgttccctattgagggcggggtgcatcaatccagagaa 243
Db 230 GTACCAAAAAACCCWATAAAGAAATGTTCCCWATTGAGGCGAGGGCGGAWCAATCCAGAGAA 289
Qy 244 caggatcatcccccacacgtccagctcccaagtgtgttgagcggcagcggcgcgcgagaa 303
Db 290 CAGGATCATCCCCACCTCCAGCTCCCAAGTGTGTGACGGGAGCGGGCCGCCGAGAA 349
Qy 304 ccgcccagggaacattaccgagtagtgcagcacgtctatgagcctgcccacaaagatctc 363
Db 350 CCGCAGGGCACACTTACCGAGTACTGCAGCACGCTCATGAGCCTGATGAGCTGCCACCAAGATCTC 409
Qy 364 ccgctgtcccccacacgtccagcttctcaaggtgcgcctgagatgaactcccccac 423
Db 410 CCGTGTGCCCCACCTCTCGACTTCTTCAAGGTGGCCCTGATGACCTCAAGCTCCCCAC 469
Qy 424 ggacaaccagacaaaaagccagagacatacttgcgtcccaagatggcgaagtagtcgc 483
Db 470 GGACAACCAAGAAAAAGCCAGAGACATCTTGCATGCCCAAAAGATGGCAAGATACCGC 529
Qy 484 gacagacatcccgcccccacatcactcgtcagacgtacgcgcccattgccaactacgagaa 543
Db 530 GACAGAAATACCGCGCCCATCATCATGTCAGACGCTACCGGCCATTGCCAACTACGAGAA 589
Qy 544 gacctcgggtccagagatggctctgcacgggggagcgtgtgagagtcgtagagaagag 603
Db 590 GACCTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCTGTGTGAGGTCTGTAGAGAAGAG 649
Qy 604 cgagagcgtgtgtgttctgtcagatgaagcaagcagcgtgtggtatcccgctcctt 663
Db 650 CGASAGGTTGGTGTCTGTCTAGATGAAGCAAGCGAGCTGGATCCAGCGCTCCTT 709
Qy 664 cctcgagcccttgacagtcgtcagcagcgaagaacctgagccccactatgcaggtga 723
Db 710 CCTCGAGCCCTTGACAGTCTCTGACGAGACGCAAGACCTCAGCCCCAATATGCAAGTGA 769
Qy 724 gceatagtcgceatcaaggctcactgctgtgagggggagcagaggtgtccctgctcga 783
Db 770 GCCATAGCTGCCATCAAGGCTACACTGTGTGTGAGGGGACGAGGTGTCCCTGCTCGA 829
Qy 784 ggggtgaagctgttgaggtcattcac 808
Db 830 GGGTGAAGCTGTGAGGTCTATTAC 854

RESULT 11
AL583335 889 bp mRNA EST 16-FEB-2001
LOCUS AL583335 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL012YL03 5
DEFINITION prime, mRNA sequence.
ACCESSION AL583335
VERSION AL583335.1 GI:12952196
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..889
/db_xref="taxon:9606"
```

```
/clone="CS0DL012YL03"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 211 a 274 c 248 g 148 t 8 others
ORIGIN

Query Match 57.0%; Score 787.8; DB 10; Length 889;
Best Local Similarity 96.8%; Pred. No. 6,1e-109;
Matches 844; Conservative 6; Mismatches 5; Indels 17; Gaps 4;

Qy 134 caggacacctgtcgagaaggtgtgtctaccggttcaccgagatctacgagttccataaa 193
Db 1 CAGGACCTGTCCGAGAAGGTGTCTACCGCGCTTACCAGATCTTACGAGTTCATATAA 60
Qy 194 accttaaaagaaatgttccctattgagcggcggcgtacatccagagaaacagatcatc 253
Db 61 ACCTTAAAGAAATGTTCCCTATTGAGGCGGGCGGATCAATCCAGAGAACAGGATCATC 120
Qy 254 cccacacctccagctcccaagtgtgttgacgggagcggcggcggcggcggcggcggcggc 313
Db 121 CCCACACTCCACAGTCCCAAGTGTGTGACGGCAGCGGGCGCGCGAGAACCGCAGGGC 180
Qy 314 acactaccgagtagctgcagcagctcatgagcgtgtccacacagatctcccgctgtccc 373
Db 181 ACATTACCGAGTAGTCTGCAGCACGCTCATGAGCCTGCCACCAAGATCTCCGCTGTCCC 240
Qy 374 cactctcgtactctcaaggtgtcgcctgtatgactcgaagctcccccagcagaacacag 433
Db 241 CACTCTCTCGACTTCTTCAAGGTGCGCCCTGTATGACCTCAAGCTCCCCACGAGCAACCG 300
Qy 434 aaaaaagccagagacatacttgcgtcccaagatggcaagagtagccgcagacagatc 493
Db 301 AAAAAAGCCAGAGACATCTTGATGCCCAAGATGGCAAGATACCGCGACACATC 360
Qy 494 accggccccatcactcgtcagacagtagccgcgcctattgccaactacgagaagacctcgggc 553
Db 361 ACCGGCCCCATCATCTCTGCAGACGTACCGCGCCATTGGCAACTACGAGAAKWTYTKRG 420
Qy 554 tccgagatggtctgtccacggggagcgtgtgaggtcgttagaagagcagagcgggt 613
Db 421 TCCGAGATGGCTCTGTCCACGGGGAGCGTGTGTGAGGTCTGTAGAGAAAGCGAGAGC 480
Qy 614 tgggtgttctgtcagatgaagcaagcagcgtgtgtggtatcccgaggtccttctcgtgagccc 673
Db 481 TGGTGGTTCTGTAGATGAAGCAAGCGAGGCTGGATGCCCGAGCTCTCTCTCGAGCCC 540
Qy 674 ctgagacgtctcgtcagacagcgaagacctgagcccaactatgcaggtgagccatcgc 733
Db 541 CTGACAGTCTCTGACGAGACGGAAGACCTCTGAGGCCAACTATGCAAGGTGAGCCATAC 600
Qy 734 gceatcaaggcctacactgtgttgagggggagcagaggtgtccctgctcgtcaggggtaagct 793
Db 601 GCCATCAAGGCTTACACTGTGTGTGAGGGGAGCGAGGTGTCCCTGCTCGAGGGTGAAGCT 660
Qy 794 gttgaggtcattcaagactccttgacgggtc-----ggaaagacgagcgtcac 841
Db 661 GTTGAGGTCTATTACAGACTCTCTGGACGGCTGTGGGTGTCATCAGGAAGAGACAGCTCACA 720
Qy 842 gggctactcccgctccatgtacctgcaaaagtccaggcgaagcgtgttccagggcccaagcgc 901
Db 721 GGTACTTCCCGTCCATGTACTGTGCAAAAVTCAGGGCAAGAGCTGTCCAGGCCCAAGGC 780
```



```
Db 62 CAGCCAGCACTATGTGTACATGTTCTGTGTAATGTCGAGGACCTGTTCGGAGAGGTGGT 121
Qy 157 ctaccgggcttcacggagatcagagttccataaaaccccttaaaagaatgttccctat 216
Db 122 CTACCGGGCTTTCACCGAGATCTAGAGTTTCATATAAACCTTAAAGAAATGTTCCCTAT 181
Qy 217 tgaggcaggggcgatcaatccagagaaacaggatcctcccccacccctccagctcccaagt 276
Db 182 TCAGGCGAGGGGATCATCCAGAGAACAGATCATCCCCACCTCCAGCTCCCAAGTG 241
Qy 277 gttgacggggaagggcgccggagaaacggcagggcacacttaccagtagtactgcagcac 336
Db 242 GTTTGACGGGACAGCGGGCGCGAGAACCGCAGGACACTTACCGAGTAGTCTGCAGCAC 301
Qy 337 gctcatgagctgcggcacaagatctccgctgttcccaacctctctgactcttccaaagt 396
Db 302 GCTCATGAGCTTCCGCCACCAAGATCTCCGCTGTCTCCACCTCTCTCGACTTCTTCAAGT 361
Qy 397 ggcctctgatgacctcaagctcccccacggacaaacccagacaaacccagagacatactt 456
Db 362 GCGCCTGATGACCTCAAGCTCCCGACGACAAACCCAGACACAAAAGCCAGACATACTT 421
Qy 457 gatgcccagaagatggaagagtagtaccgagacagacatcccgcccccacatcctctgcagac 516
Db 422 GATGCCCAAGATGCAAGATACCGGACAGACATCACCGGCCCATCATCTCTGCAGAC 481
Qy 517 gtaccgcccattgccaactacgagaaagacctcgggctccagatggctctgtccacgg 576
Db 482 GTACCGGCCCATTGCCACTACGAGAAAGCTTCGGGCTCCCGAGATGGCTCTGTCCACGGG 541
Qy 577 ggaactggtgaggt-cgtagagaagagcagagcgggtggtggtctgtcagatgaag 635
Db 542 GCACCTGTGGAGTCCGTAGAGAGAGCGAGAGCGGTTGCTGTCTGTCTGATGAAG 601
Qy 636 caaagcagggtgtagtccagcagctctctctcagacccctggaagctcctcagaag-9acg 694
Db 602 CAAAGCGAGGTGGATCCAGCGCTTTCTCCGAGGCCCTGGACAGTCTCTCAGCAGGACG 661
Qy 695 gaagacctgagcccaact--atgaggtgagccatagc-tcgccatcaagcctacact 751
Db 662 GAAGACCTTGAGCCCACTTATGAGGGTGAGCCATAGCTTTCGCCATCAAGGCTTACACT 721
Qy 752 gctgtgagggggacgaggttccctgctcagaggtgaggtgaggtgttggagcttaccacaa 811
Db 722 GCTGTGTAGGGGAACGAGGTGTCCTCTGCTCGAGGTGAAGCTGTTGAAGTCAATTTCACAG 781
Qy 812 ctcttgagcg 821
Db 782 CTCTGGGACG 791

RESULT 15
AL581347/c
LOCUS AL581347 LTI_FL011_Bcl Homo sapiens cDNA clone CS0DG005YE05 3 prime
DEFINITION , mRNA sequence.
ACCESSION AL581347
VERSION AL581347.1 GI:12948254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
source
1. .834
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG005YE05"
/lab_host="DHL10B"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DHL10B"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen. 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 126 a 238 c 271 g 191 t 8 others
ORIGIN
Query Match 51.5%; Score 712.2; DB 10; Length 834;
Best Local Similarity 95.0%; Pred. No. 1.3e-97;
Matches 777; Conservative 5; Mismatches 19; Indels 17; Gaps 4;
Qy 318 ttaccgagtagtcagcagcgtcatgagcctgcccaccaaagatctccgctgtcccaacc 377
Db 816 TTACGGAGTACTGCAS-ACGCTCATGAGCTTGCCACCAAGATCTCCGCTGTCCCAACC 758
Qy 378 tctctgactcttcaaggtgcgcctgatgacctcaagctcccccacggacacacagacaa 437
Db 757 TCCTCGACTTCTCAAGGTGCGCCTGATGACCTCAAGCTCCCGACGACACACAGAC-A 699
Qy 438 aaaagcagagacatacttgatgccaaaagatgacagagtagtaccgagacagacatccacg 497
Db 698 AAAAGCCAGAGACATACTTGTATGCCAAAGATGGCAAGTAGTACCGCGAGACATCACCG 639
Qy 498 gcccatcatctcagacgtaccgcgcattgccaactacgagaagacctcgagctccg 557
Db 638 GCCCATCATCTCTCGAGACGTACCGGCCCATTCGCGGACTACGAGAACACCTTCGGGCTCCG 579
Qy 558 agatggtctgttccacgggggagcgtggtgaggtcgtagaagagcagagcgtggttgg 617
Db 578 AGATGGGTCTCTCCACGGGGACGCTGTGTGAGGTCTGTGAGAGAGAGAGAGAGCGTGGT 519
Qy 618 ggttctgcaagataaaagcagaggtgtagtcccaagtcctctctcagagcctcctg 677
Db 518 GGTCTGTGATGAAAGCAAGAGGAGCTGTGATCCCGAGCATCTCTCTCGAGGCCCTGG 459
Qy 678 acagtccgtgacgagcaggaagacccctgagcccaactatgaggtgagcctacacgcca 737
Db 458 ACAGTCTCTGAGAGCGAGAGACCCCTGAGCCCACTATGCAAGGTGAGCCATACGTCGCCA 399
Qy 738 tcaaggctacactgctgtgaggggagcaggtgttccctgctcagaggtgtgaagctgtt 797
Db 398 TCAAGGGCTACACTGCTGTGAGGGGACGAGGTGCTCCTCTCGAGGKTGAAGCTGTG 339
Qy 798 aggtcattcaagctcctgagcgtctgg-----aaagcagcgtcacagggc 844
Db 338 AGKTCATTCACAAGCTCTTGAGCGGCTGGTGGTTCATCAGGAAAGAACATACACAGGC 279
Qy 845 tacttccgctcatgtacctgcaaaagtccaggcgaagacgtgtcccgagcccaacgcag 904
Db 278 TACTTTCCGTCCTCATGTACCTTGCAAAAGTCGGGGCAAGACGTGTCCAGGCCCAAGCCAG 219
Qy 905 atcaagcggggg--gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 962
Db 218 ATCAAGCGGTGTGTGTCGCGCGCGCGCGCGAGGTGTCATCCGCAACGCGCACAAATCATC 159
Qy 963 agcgttcgcggaagcgtcctcagcagcagccttatcccccacacagcgtcgttttttc 1022
Db 158 AGCGGTGCGGGAAGCGCTCAGCCAGGACSCCTTATCCCGGCAACAGCGCTCCGTTTCTGC 99
Qy 1023 agcagcagcgcgcgcagggcgcgccgggacgcagagccccgggagccccgctcagaggg 1082
```

|||||
Db 98 AGCAGCGACCCNCCAGGCGCGCCGGACCGCAGAGCCCGGGAGCCCGCTCGAGGAGG 39
Qy 1083 agcggcagacgcagcgtctctaaaccgcagccgcggig 1120
|||
Db 38 AGCWGACAGCCAGCGCTCTAAACCGCAGCGCGCGGTG 1

Search completed: January 11, 2002, 11:34:22
Job time: 8814 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 09:07:28 ; Search time 814.02 Seconds
(without alignments)
1455.521 Million cell updates/sec

Title: US-09-820-005-1
Perfect score: 1382
Sequence: 1 cctggaagtcgacggagca.....aaaaaaaaaaaaaaaa 1382

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	384	27.8	384	AAH45277	DNA encoding human
2	323.8	23.4	425	AAC98438	Human colon cancer
3	154.8	11.2	192	AAH45280	DNA encoding human
4	99.8	7.2	3195	AA522715	Human cDNA encoding
5	88	6.4	5588	AAS22479	Human cDNA encoding
6	77.2	5.6	7368	AA522479	Human low adenodin
7	77.2	5.6	11651	AA522479	Human low adenodin
8	75	5.4	75	AA522479	Human low adenodin
9	75	5.4	75	AA522479	Human low adenodin
10	52.4	3.8	3201	AA522479	DNA encoding a PDE
11	51.6	3.7	1214	AAC99095	Human pancreatic c

c	12	49	3.5	1305	22	AA507945	Human CDNA encodin
c	13	49	3.5	1305	22	AAH73513	Human G protein-co
c	14	49	3.5	1316	21	AAH75192	Human OREX ORF747
c	15	48.6	3.5	114955	20	AA533491	Human adenosine A1
	16	47.8	3.5	1803	20	AA589760	Nucleotide sequenc
	17	47.8	3.5	1850	21	AA246151	cDNA sequence enco
	18	47.8	3.5	1858	20	AA589759	Nucleotide sequenc
	19	47.8	3.5	1974	21	AA589759	Human OREX ORF2727
	20	46.4	3.4	1635	21	AA589759	Human secreted pro
	21	46.4	3.4	2115	20	AA589759	Human dunce-like p
	22	46.4	3.4	2115	20	AA589759	ppD39 human dunce
	23	46.4	3.4	2205	22	AA589759	Human secreted pro
	24	46	3.3	734	19	AA589759	Human secreted pro
	25	46	3.3	1513	11	AA589759	Chicken ovalbumin
	26	45.8	3.3	570	21	AA589759	Membrane-bound pro
	27	45.8	3.3	570	22	AA589759	Human PRO1245 (UNQ
	28	45.8	3.3	1163	14	AA589759	ppCDHP71 coding se
	29	45.8	3.3	1541	14	AA589759	ppCDHP23 coding se
	30	45.6	3.3	1043	21	AA589759	cDNA encoding huma
	31	45.6	3.3	1050	21	AA589759	Human secreted pro
	32	45.6	3.3	1050	22	AA589759	Human secreted pro
	33	45	3.3	1233	21	AA589759	Human breast and o
	34	44.6	3.2	10732	21	AA589759	Gene encoding a su
	35	44.4	3.2	2004	17	AA589759	Maize-optimized VI
	36	44.4	3.2	2004	18	AA589759	Maize-optimized-B.
	37	44.4	3.2	2004	19	AA589759	DNA encoding an 80
	38	44.4	3.2	2010	15	AA589759	Bacillus cereus ve
	39	44.4	3.2	2576	17	AA589759	Maize-optimized VI
	40	44.4	3.2	2576	18	AA589759	Maize-optimized-B.
	41	44.4	3.2	2576	19	AA589759	Maize-optimized DN
	42	44.4	3.2	2655	15	AA589759	Bacillus cereus ve
	43	44.4	3.2	2655	17	AA589759	Maize-optimized VI
	44	44.4	3.2	2655	17	AA589759	Maize-optimized VI
	45	44.4	3.2	2655	18	AA589759	Maize-optimized-B.

ALIGNMENTS

RESULT 1
AAH45277
ID AAH45277 standard; DNA; 384 BP.
XX
AC AAH45277:
XX
DT 07-SEP-2001 (first entry)
XX
DE DNA encoding human PX domain.
XX
KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH;
KW NADPH oxidase; superoxide production; inflammatory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200142453-A1.
XX
PD 14-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-JP08501.
XX
PR 06-DEC-1999; 99JP-0346193.
XX
PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX
PI Kohda D, Hiroaki H, Sumimoto H;
XX
DR WPI; 2001-381679/40.
XX
PT Regulation of function of a protein containing a PX domain for
PT controlling an inflammatory response to disease or trauma, comprises
PT using nuclear magnetic resonance (NMR) chemical shift and structural

PT coordinate data -

XX Example 1; Page 185-186; 195pp; Japanese.

XX The present sequence is provided in a specification relating to the use
 CC of chemical shift data from nuclear magnetic resonance (NMR) and
 CC structural coordinate data of the PX domain of a protein to search for,
 CC evaluate, design and identify variant PX domain sequences for
 CC controlling the function of proteins containing the PX domain. It also
 CC relates to compounds promoting the binding of substances to the PX
 CC domain, and compounds inhibiting the binding of substances to the PX
 CC domain. The invention can be used in the identification of substances
 CC that can be used in the regulation of reduced nicotinamide adenine
 CC dinucleotide phosphate (NADPH) oxidase activity, which is involved in
 CC superoxide production in inflammatory response in disease or trauma.

XX Sequence 384 BP; 86 A; 127 C; 95 G; 76 T; 0 other;

Query Match 27.8%; Score 384; DB 22; Length 384;
 Best Local Similarity 100.0%; Pred. No. 3.8e-72;
 Matches 384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 atggggagacacattcctcgcacatcgccctgctgggttgagaagcgttcgtaccc 97
 Db 1 atggggagacacattcctcgcacatcgccctgctgggttgagaagcgttcgtaccc 60
 Qy 98 agccagacactatgtacatgttctctgtgaaatggcaggacctgtcggaagtgtgc 157
 Db 61 agccagacactatgtacatgttctctgtgaaatggcaggacctgtcggaagtgtgc 120
 Qy 158 taccggcgcttcacagatcacgtacagttccataaaaccttaaaagaaatgttcctatt 217
 Db 121 taccggcgcttcacagatcacgtacagttccataaaaccttaaaagaaatgttcctatt 180
 Qy 218 gaggcaggggcgatcaatccagagaacagagatcatccccacctcccagctcccaagtgg 277
 Db 181 gaggcaggggcgatcaatccagagaacagagatcatccccacctcccagctcccaagtgg 240
 Qy 278 tttagcgggagcggcgccgagaaacccagggacacttaccgagtactgcagcag 337
 Db 241 tttagcgggagcggcgccgagaaacccagggacacttaccgagtactgcagcag 300
 Qy 338 ctcatgagcctgcccacacagatctccgctgtccacacctcttcgacttttcaagtgg 397
 Db 301 ctcatgagcctgcccacacagatctccgctgtccacacctcttcgacttttcaagtgg 360
 Qy 398 cgcctgatgacctcaagctcccc 421
 Db 361 cgcctgatgacctcaagctcccc 384

RESULT 2

AAC98438
 ID AAC98438 standard; cDNA; 425 BP.

XX AAC98438;

XX 09-MAR-2001 (first entry)

XX Human colon cancer antigen nucleotide sequence SEQ ID NO:448.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antinefactive; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder; ss.

XX Homo sapiens.

XX WO200055351-A1.

Qy 464 aaagatggcaagagtlaccgagacagatcacccggcccccattcatctcgtgcagcgtacccg 523

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587534/55.

XX P-PSDB; AAB53681.

XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

XX Claim 1; Page 930; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnery, nephrotropic, antinefactive and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 425 BP; 103 A; 145 C; 102 G; 75 T; 0 other;

Query Match 23.4%; Score 323.8; DB 21; Length 425;
 Best Local Similarity 84.1%; Pred. No. 1.9e-59;
 Matches 423; Conservative 0; Mismatches 2; Indels 78; Gaps 2;

Qy 44 gacaccttcacgtcaccatcgccctgctgggttgagaagcgttcgtaccagccag 103

Db 1 gacaccttcacgtcaccatcgccctgctgggttgagaagcgttcgtaccagccag 60

Qy 104 cactatgtgtacatgttctctgtgaaatggcaggacctgtcggaaggtgtgtaccgg 163

Db 61 cacta--tgtacatgttctctgtgaaatggcaggacctgtcggaaggtgtgtaccgg 118

Qy 164 cgttccaccagatctacgagttccataaaaccttaaaagaaatgttccctattgagcca 223

Db 119 cgttccaccagatctacgagttccat-----ctcccaagtggtttgac 145

Qy 224 ggggcgatcaatccagagaacagagatcatcccccacgtccacagctcccaagtggtttgac 283

Db 146 -----ctcccaagtggtttgac 162

Qy 284 gggcagcggcgcccgagaaacccgagggcacacttaccgagtactgcagcagctcatg 343

Db 163 gggcagcggcgcccgagaaacccgagggcacacttaccgagtactgcggaagctcatg 222

Qy 344 agcctgccaccaagatctccgctgtcccccactctcgtacttctcaaggtgcgcct 403

Db 223 agcctgccaccaagatctccgctgtcccccactctcgtacttctcaaggtgcgcct 282

Qy 404 gatgacctcaagctccccacggagcaaccagacaaagagccagagacatacttattgccc 463

Db 283 gatgacctcaagctccccacggagcaaccagacaaagagccagagacatacttattgccc 342


```
Db 343 aaagatgcaagatgacgcgacagacatcacccggcccatcctcctgcagagctaccgc 402
QY 524 gccattgccactacgagaagac 546
Db 403 gccattgccactacgagaagac 425

RESULT 3
AAH45280
ID AAH45280 standard; DNA: 192 BP.
XX AC AAH45280;
XX DT 07-SEP-2001 (first entry)
XX DE DNA encoding human SH3 domain.
XX KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH; SH3 domain;
KW NADPH oxidase; superoxide production; inflammatory disease; ds.
XX OS Homo sapiens.
XX PN WO20012453-A1.
XX PD 14-JUN-2001.
XX PF 01-DEC-2000; 2000WO-JP08501.
XX PR 06-DEC-1999; 99JP-0346193.
XX PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX PI Kohda D, Hiroaki H, Sumimoto H;
XX DR WPI; 2001-381679/40.
XX PT Regulation of function of a protein containing a PX domain for
XX controlling an inflammatory response to disease or trauma, comprises
XX using nuclear magnetic resonance (NMR) chemical shift and structural
XX coordinate data.
XX PS Example 2; Page 186; 195pp; Japanese.
XX CC The present sequence is provided in a specification relating to the use
XX of chemical shift data from nuclear magnetic resonance (NMR) and
XX structural coordinate data of the PX domain of a protein to search for,
XX evaluate, design and identify variant PX domain sequences for
XX controlling the function of proteins containing the PX domain. It also
XX relates to compounds promoting the binding of substances to the PX
XX domain, and compounds inhibiting the binding of substances to the PX
XX domain. The invention can be used in the identification of substances
XX that can be used in the regulation of reduced nicotinamide adenine
XX dinucleotide phosphate (NADPH) oxidase activity, which is involved in
XX superoxide production in inflammatory response in disease or trauma.
XX SQ Sequence 192 BP; 44 A; 50 C; 61 G; 37 T; 0 other;

Query Match 11.2%; Score 154.8; DB 22; Length 192;
Best Local Similarity 92.1%; Pred. No. 6.8e-24;
Matches 178; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 704 gagcccaactatgcagtgagccatcgtcgccatcaagccctacactgctgtggaagg 763
Db 1 gagcccaactatgcagtgagccatcgtcgccatcaagccctacactgctgtggaagg 60
QY 764 gagagtggtccctgctcgaggggtgaagctgttgaggtcattcacaagctccctggagcgc 823
Db 61 gagagtggtccctgctcgaggggtgaagctgttgaggtcattcacaagctccctggagcgc 120
```

```
QY 824 t-----ggaaagacgacgtcacagagctactccctccatctactctgcaaaag 871
Db 121 tgggtggtcatcagaaaagacgacgtcacagagctacttccgtccatctgtaaaag 180
QY 872 tcaggggcaagac 883
Db 181 tcgggggcaagac 192

RESULT 4
AAS22715
ID AAS22715 standard; cDNA: 3195 BP.
XX AC AAS22715;
XX DT 24-OCT-2001 (first entry)
XX DE Human cDNA encoding a novel human protein #281.
XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX OS Homo sapiens.
XX PN WO200155437-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02623.
XX PR 25-JAN-2000; 2000US-0491404.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-451939/48.
XX DR P-PSDB; AAU14410.
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage.
XX CC Claim 1; Page 694-695; 894pp; English.
XX CC The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/elicite an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
```


KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; ROS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

XX WO2000062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions.

XX Disclosure; Page 1383-1385; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 7368 BP: 2018 A; 1656 C; 1890 G; 1804 T; 0 other;

Query Match 5.6%; Score 77.2; DB 21; Length 7368;
 Best Local Similarity 96.3%; Pred. No. 3.4e-07;
 Matches 79; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 50 ttctcgtcacatcgccctgcttgaagcgttcgtaccagccagcactat 109

Db 7287 ttctgaagtcacatcgccctgcttgaagcgttcgtaccagccagcactat 7346

Oy 110 gtgtacatgttctcgtggaat 131
 |||||||

Db 7347 gtgtacatgttctcgtggaat 7368

RESULT 7

AAA35276

ID AAA35276 standard; DNA; 11651 BP.

XX AC AAA35276;

XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:150.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers.
 XX Disclosure; Page 1298-1301; 1343pp; English.

XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA3512 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX Sequence 11651 BP; 3079 A; 2635 C; 2989 G; 2948 T; 0 other;

Query Match	5.6%;	Score 77.2;	DB 21;	Length 11651;
Best Local Similarity	96.3%;	Pred. No. 3,7e-07;		
Matches	79;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	50	ttcatcgtcacatcgccctgctgggttttgagaagcgcttcgtaccacgacgacactat 109		
Db	11570	ttctaaagtcacatcgccctgctgggttttgagaagcgcttcgtaccacgacgacactat 11629		
QY	110	gtgtacatgttcctcgttgaaat 131		
Db	11630	gtgtacatgttcctcgttgaaat 11651		
RESULT	8			
ID	AAF21397	AAF21397 standard; DNA; 75 BP.		
AC	AAF21397;			
XX	AAF21397;			
XX	14-MAR-2001	(first entry)		
XX	Human low adenosine antisense oligonucleotide related sequence #2964.			
DE				
XX				
KW	Low adenosine antisense oligonucleotide; phosphorothioate; allergy;			
KW	human; airway disorder; bronchoconstriction; lung inflammation;			
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;			
KW	immunosuppressive; antialsthmatic; analgesic; hypotensive; cytostatic;			
KW	respiratory obstruction; pulmonary obstruction; impeded respiration;			
KW	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;			
KW	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;			
KW	pulmonary hypertension; emphysema; pulmonary transplantation rejection;			
KW	chronic obstructive pulmonary disease; pulmonary infection; bronchitis;			
KW	cancer; ss.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200062736-A2.			
XX				
XX	26-OCT-2000.			
PD				
XX				
XX	24-MAR-2000; 2000WO-US08020.			
PF				
XX				
PR	06-APR-1999; 99US-0127958.			
XX				
PA	(UYEC-) UNIV EAST CAROLINA.			
PA	(NYCE/) NYCE J W.			
XX				
PI	Nyce JW;			
XX				
DR	WPI; 2000-679539/66.			
XX				
XX	Low adenosine (A) content antisense oligonucleotides which do not			
PT	trigger adenosine receptors during metabolism, useful e.g. for treating			
PT	cancers and respiratory obstructions -			
XX				
PS	Disclosure; Page 1383; 1592pp; English.			
XX				
CC	The present invention describes low adenosine (A) content antisense			
CC	oligonucleotides and compositions (I) comprising them. In the antisense			
CC	oligonucleotides the A is replaced by a 'Universal' or alternative base.			
CC	(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,			
CC	immunosuppressive, antialsthmatic, hypotensive and cytostatic activities.			
CC	The antisense oligonucleotides and (I) can be used to down-regulate the			
CC	expression and or activity of target polypeptides associated with			
CC	lung/respiratory disorders and malignancies, such as stimulating and			
CC	activating peptide factors and transmitters, transcription factors,			
CC	immunoglobulins and antibodies, antibody receptors, cytokines and			
CC	chemokines, endogenously produced specific and non-specific enzymes,			
CC	binding proteins, adhesion molecules and their receptors, cytokine and			
CC	chemokine receptors, adenosine receptors, bradykinin receptors, central			
CC	nervous system (CNS) and peripheral nervous and non-nervous system			
CC	receptors, CNS and peripheral nervous and non-nervous system peptide			

AAC99095
ID AAC99095 standard; cDNA; 1214 BP.
AC AAC99095;
XX
DT 09-MAR-2001 (first entry)
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:323;
XX
XX
KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
DR P-PSDB; AAB54330.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1; Page 748; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 1214 BP; 264 A; 395 C; 325 G; 229 T; 1 other;

Query Match 3.7%; Score 51.6; DB 21; Length 1214;
Best Local Similarity 58.4%; Pred. No. 0.057;
Matches 90; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 1229 ggccttcgcccgcgccttcgtatctacgtgttcctatagacgtgcgttcgacgcgc 1288
Db 1040 gtcacttcgctgccttcgttcgtgtgtgagcagagtgagctggggggcg 1099

Qy 1289 agggcagcccgacccctgtccagcggtccgcgacccctcaataatgttcttgga 1348
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1100 tggagagcccgccgcccctccctccctccctgacccgtctatataatagagatgt 1159
Qy 1349 gtggaaaaaataaaaaaataaaaaaataaaaaa 1382
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1160 gtctaaaaaataaaaaaataaaaaaataaaaaa 1193

RESULT 12
AAS07945/c
ID AAS07945 standard; cDNA; 1305 BP.
XX
AC AAS07945;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human cDNA encoding G-protein coupled receptor, hRUP18.
XX
KW Human; G-protein coupled receptor; GPCR; hRUP18; agonist;
KW inverse agonist; lung cancer; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1305
FT FT /*tag= a
FT FT /product= "hRUP18"
XX
PN WO200136471-A2.
XX
PD 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31509.
XX 17-NOV-1999; 99US-0166088.
XX 17-NOV-1999; 99US-0166099.
XX 17-NOV-1999; 99US-0166369.
XX 23-DEC-1999; 99US-0171900.
XX 23-DEC-1999; 99US-0171901.
XX 23-DEC-1999; 99US-0171902.
XX 11-FEB-2000; 2000US-0181749.
XX 14-MAR-2000; 2000US-0189258.
XX 14-MAR-2000; 2000US-0189259.
XX 10-APR-2000; 2000US-0195898.
XX 10-APR-2000; 2000US-0195899.
XX 10-APR-2000; 2000US-0196078.
XX 28-APR-2000; 2000US-0200419.
XX 12-MAY-2000; 2000US-0203630.
XX 12-JUN-2000; 2000US-0210741.
XX 12-JUN-2000; 2000US-0210982.
XX 21-AUG-2000; 2000US-0226760.
XX 26-SEP-2000; 2000US-0235418.
XX 26-SEP-2000; 2000US-0235779.
XX 20-OCT-2000; 2000US-0242332.
XX 20-OCT-2000; 2000US-0242343.
XX (AREN-) ARENA PHARM INC.
XX Chen R, Dang HT, Lowitz KP;
PI P-PSDB; AAU04372.
XX WPI; 2001-355616/37.
DR P-PSDB; AAU04372.
XX
XX Endogenous and non-endogenous versions of human G-protein coupled
XX receptors for direct identification of candidate compounds as agonists,
XX inverse agonists or partial agonists for use as therapeutic agents -
XX
XX Claim 43; Page 108; 159pp; English.
XX The sequence encodes a human G-protein coupled receptor (GPCR),
XX hRUP18 The endogenous and non-endogenous, constitutively activated
CC

versions of human G-protein coupled receptors (GPCR), are useful for direct identification of candidate compounds as receptor agonists, inverse agonists or partial agonists having applicability as therapeutic agents for treating diseases related to GPCR, e.g. lung cancer. Non-endogenous version of human GPCRs are also utilized in research settings and in vitro and in vivo system, incorporating GPCRs can be utilised to elucidate and understand the roles these receptors play in the human condition, both normal and diseased.

Sequence 1305 BP; 127 A; 500 C; 467 G; 211 T; 0 other;
XX
SQ

SQ Sequence 1305 BP; 127 A; 500 C; 467 G; 211 T; 0 other;

Query Match 3.5%; Score 49; DB 22; Length 1305;
Best Local Similarity 50.5%; Pred. No. 0.2;
Matches 147; Conservative 0; Mismatches 140; Indels

[illegible]

RESULT 13
AAH73513/C
ID AAH73513 standard; DNA: 1305 BP.

XX	AAH73513;
AC	
XX	
DT	25-SEP-2001 (first entry)
XX	
XX	Human G protein-coupled receptor GPRv40 coding sequence.
DE	
XX	Human; guanosine triphosphate binding protein-coupled receptor;
KW	G protein-coupled receptor; GPRv8; GPRv12; GPRv16; GPRv21; GPRv40
KW	GPRv47; GPRv51; GPRv71; GPRv72; cancer; liver cirrhosis;
KW	Alzheimer's disease; cytostatic; hepatotrophic; nootropic;
KW	neuroprotective; gene therapy; peptide therapy; ds.

XX OS Homo sapiens.
XX PN WO200148188-A1.
XX PD 05-JUL-2001.
XX PF 28-DEC-2000; 2000WO-JP09408.
XX PR 28-DEC-1999; 99JP-0375152.
XX PR 31-MAR-2000; 2000JP-0101339.
XX PA (HELI-) HELIX RES. INST.

XX Matsumoto S, Oda T, Saito Y, Morikawa N, Yoshida K, Suwa M;
 PI Sugiyama T, Kishimoto T, Kanzaki K, Yasuda S, Inoue Y;
 XX WPI; 2001-425662/45.
 DR P-PSDB: AAG64122.

PT New DNA encoding quanosine triphosphate binding protein coupled

PT receptors and their expression products for screening potential
PT anticancer and neurotropic drugs and in diagnosis of these diseases -
XX
PS Claim 1; Page 139-140; 170pp; Japanese.
XX
CC The invention relates to nine human guanosine triphosphate binding
CC protein (G protein)-coupled receptors designated GPRv8, GPRv16,
CC GPRv21, GPRv40, GPRv47, GPRv51, GPRv71 and GPRv72, and to the
CC genes encoding them. These genes and proteins and antibodies against
CC the protein are useful in the treatment, prevention, diagnosis and
CC investigation of diseases associated with G protein-coupled receptors,
CC including cancer, cirrhosis of the liver and Alzheimer's disease.
CC The present sequence is a G protein-coupled receptor coding sequence
CC of the invention.
XX
SO Sequence 1305 BP; 127 A; 501 C; 466 G; 211 T; 0 other;

Query Match 3.5%; Score 49; DB 22; Length 1305;
Best Local Similarity 50.5%; Pred. No. 0.2;
Matches 147; Conservative 0; Mismatches 140; Indels 4; Cons

[illegible]

RESULT 14
AAC75192
ID AAC75192 standard; cDNA; 1316 BP.
XX
XX AAC75192;
XX AC
XX DT
XX 08-FEB-2001 (first entry)
XX
XX Human ORF ORF747 polynucleotide sequence
DE DE Human ORF ORF747 polynucleotide sequence

Human; open reading frame; ORF; detection; cytosolic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinfluenza; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive; ss.

AA	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:04:08 ; Search time 814.02 Seconds
(without alignments)
3475.557 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300
Perfect score: 3300
Sequence: 1 tactaaaaatacaaaattag.....gtctaggattacaacgctg 3300

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c 1	358.4	10.9	32152	22	AAI57791 Human colorectal c
c 2	356.4	10.8	32204	22	AAI57790 Human colorectal c
c 3	339.8	10.3	10503	22	AAI63970 Human polynucleoti
c 4	337.4	10.2	39198	22	AAF58067 Human polyamine-mo
c 5	333.2	10.1	13864	22	AAF97861 Human neuroblastom
c 6	332.6	10.1	20444	22	AAI62640 Human breast or ov
c 7	329.8	10.0	1334	21	AAAB1655 N. meningitidis pa
c 8	326.6	9.9	121162	21	AAAC66548 Human kinesin-like
c 9	324	9.8	7588	22	AAI62714 Human breast or ov
c 10	323.8	9.8	21721	20	AAH83427 Human lipolysis st
c 11	323.8	9.8	22976	20	AAH83426 Genomic region con

c 12	323.8	9.8	23187	21	AAA50273 Human lipolysis st
c 13	323.8	9.8	23187	22	AAF62331 Human leptin fragm
c 14	323	9.8	240825	22	AAF24497 Human PG-3 gene.
c 15	320.6	9.7	21278	21	AAH55965 Human G713 3'-end
c 16	320	9.7	39198	22	AAF58067 Human polyamine-mo
c 17	318.6	9.7	7146	19	AAV38933 Nucleic acid seque
c 18	318.6	9.7	160552	22	AAO20697 Human glycosyl sul
c 19	317.6	9.6	65921	21	AAH89046 Human nibrin DNA.
c 20	317.4	9.6	52216	22	AAH28355 Nucleotide sequenc
c 21	316.4	9.6	162450	21	AAH286967 Retinoblastoma bin
c 22	316.2	9.6	1162	20	AAH286967 Human secreted pro
c 23	315.4	9.6	53526	19	AAH94101 Human PKD1 gene.
c 24	315.4	9.6	53577	17	AAH18551 Human polycystic k
c 25	315.4	9.6	53577	19	AAH94108 Human PKD1 locus b
c 26	315.2	9.6	26928	20	AAH32184 Human prothrombin
c 27	314.8	9.5	12477	22	AAI63967 Human polynucleoti
c 28	314.6	9.5	51474	22	AAH97846 Human neuroblastom
c 29	314.4	9.5	2562	22	AAH17646 Human cDNA sequenc
c 30	313.4	9.5	32192	22	AAI63523 Human kidney relat
c 31	312.8	9.5	32134	22	AAI63522 Human kidney relat
c 32	312.4	9.5	15848	20	AAH32190 Human heparin cofa
c 33	312	9.5	13161	22	AAH97867 Human neuroblastom
c 34	312	9.5	25423	22	AAI57656 Human colorectal c
c 35	312	9.5	25424	22	AAI57657 Human colorectal c
c 36	311.4	9.4	45546	20	AAH23520 Human kidney amino
c 37	311.2	9.4	84607	20	AAH90847 Human PACAP genom
c 38	310.6	9.4	13161	22	AAH97867 Human neuroblastom
c 39	310	9.4	5846	22	AAH18668 Human cDNA sequenc
c 40	309.8	9.4	32204	22	AAI57790 Human colorectal c
c 41	309.8	9.4	32367	19	AAH35620 Human SHOX (short
c 42	309.4	9.4	579	22	AAH09228 Human cDNA clone f
c 43	309.4	9.4	16951	22	AAH48620 Human fascin DNA f
c 44	309.4	9.4	16951	22	AAH48622 Human fascin DNA f
c 45	309.2	9.4	4693	22	AAH18738 Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAI57791/c
ID : AAI57791 standard; DNA: 32152 BP.
AC AAI57791;
XX
XX
XX
DF 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 328.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO20015350-A1.
XX
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
PP 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-02321968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA
XX
XX
PI
XX
XX
DR
XX
XX
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
XX
XX
SQ

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-457727/49.

Isolated polypeptide for treating, preventing and/or prognosing disorders related to the colon and rectum including colorectal cancers and also for testing and detection e.g. diagnosis

Disclosure; SEQ ID NO: 328; 522pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of colorectal cancer antigens. These are shown in AA157547-AA157619 and AA38569-AA38641. These can be used in the diagnosis, prevention and treatment of cancer of the colon and/or rectum. The present sequence is a colorectal cancer antigen genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 32152 BP; 8332 A; 7335 C; 7256 G; 9229 T; 0 other;

PR	19-MAY-2000	20000US-02035417	20000US-02035417
PR	19-JUN-2000	20000US-02035417	20000US-02035417
PR	28-JUN-2000	20000US-02114886	20000US-02114886
PR	30-JUN-2000	20000US-02115135	20000US-02115135
PR	07-JUL-2000	20000US-02166477	20000US-02166477
PR	07-JUL-2000	20000US-02168800	20000US-02168800
PR	11-JUL-2000	20000US-02117487	20000US-02117487
PR	11-JUL-2000	20000US-02117496	20000US-02117496
PR	14-JUL-2000	20000US-02182900	20000US-02182900
PR	26-JUL-2000	20000US-02209633	20000US-02209633
PR	26-JUL-2000	20000US-02209643	20000US-02209643
PR	14-AUG-2000	20000US-02245118	20000US-02245118
PR	14-AUG-2000	20000US-02245119	20000US-02245119
PR	14-AUG-2000	20000US-02252113	20000US-02252113
PR	14-AUG-2000	20000US-02252114	20000US-02252114
PR	14-AUG-2000	20000US-02252266	20000US-02252266
PR	14-AUG-2000	20000US-02252267	20000US-02252267
PR	14-AUG-2000	20000US-02252268	20000US-02252268
PR	14-AUG-2000	20000US-02252270	20000US-02252270
PR	14-AUG-2000	20000US-02253474	20000US-02253474
PR	14-AUG-2000	20000US-02257577	20000US-02257577
PR	14-AUG-2000	20000US-02257578	20000US-02257578
PR	14-AUG-2000	20000US-02257579	20000US-02257579
PR	18-AUG-2000	20000US-02262679	20000US-02262679
PR	22-AUG-2000	20000US-02266681	20000US-02266681
PR	22-AUG-2000	20000US-02266688	20000US-02266688
PR	23-AUG-2000	20000US-02271822	20000US-02271822
PR	30-AUG-2000	20000US-02277009	20000US-02277009
PR	30-AUG-2000	20000US-02289924	20000US-02289924
PR	01-SEP-2000	20000US-02292877	20000US-02292877
PR	01-SEP-2000	20000US-02293433	20000US-02293433
PR	01-SEP-2000	20000US-02293444	20000US-02293444
PR	01-SEP-2000	20000US-02293445	20000US-02293445
PR	05-SEP-2000	20000US-02295909	20000US-02295909
PR	05-SEP-2000	20000US-02295913	20000US-02295913
PR	06-SEP-2000	20000US-02304737	20000US-02304737
PR	06-SEP-2000	20000US-02304738	20000US-02304738
PR	08-SEP-2000	20000US-02312422	20000US-02312422
PR	08-SEP-2000	20000US-02312423	20000US-02312423
PR	08-SEP-2000	20000US-02312444	20000US-02312444
PR	08-SEP-2000	20000US-02314113	20000US-02314113
PR	08-SEP-2000	20000US-02314114	20000US-02314114
PR	08-SEP-2000	20000US-02320800	20000US-02320800
PR	12-SEP-2000	20000US-02320801	20000US-02320801
PR	12-SEP-2000	20000US-02319668	20000US-02319668
PR	14-SEP-2000	20000US-02323977	20000US-02323977
PR	14-SEP-2000	20000US-02323988	20000US-02323988
PR	14-SEP-2000	20000US-02323999	20000US-02323999
PR	14-SEP-2000	20000US-02324000	20000US-02324000
PR	14-SEP-2000	20000US-02324011	20000US-02324011
PR	14-SEP-2000	20000US-02324063	20000US-02324063
PR	14-SEP-2000	20000US-02324064	20000US-02324064
PR	21-SEP-2000	20000US-02330655	20000US-02330655
PR	21-SEP-2000	20000US-02334223	20000US-02334223
PR	21-SEP-2000	20000US-02344274	20000US-02344274
PR	25-SEP-2000	20000US-02349977	20000US-02349977
PR	25-SEP-2000	20000US-02349998	20000US-02349998
PR	26-SEP-2000	20000US-02334484	20000US-02334484
PR	27-SEP-2000	20000US-02356834	20000US-02356834
PR	27-SEP-2000	20000US-02356836	20000US-02356836
PR	29-SEP-2000	20000US-02363627	20000US-02363627
PR	29-SEP-2000	20000US-02363667	20000US-02363667
PR	29-SEP-2000	20000US-02363668	20000US-02363668
PR	29-SEP-2000	20000US-02363699	20000US-02363699

20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0241617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-457727/49.
Isolated polypeptide for treating, preventing and/ or prognosing
disorders related to the colon and rectum including colorectal cancers
and also for testing and detection e.g. diagnosis -
Disclosure; SEQ ID NO: 327; 522pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of colorectal cancer antigens. These are shown in
AA157547-AA157619 and AAM38369-AAM38641. These can be used in the
diagnosis, prevention and treatment of cancer of the colon and/or rectum.

CC The present sequence is a colorectal cancer antigen genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32204 BP; 8487 A; 7076 C; 7452 G; 9189 T; 0 other;

Query Match 10.8%; Score 356.4; DB 22; Length 32204;
Best Local Similarity 77.6%; Pred. No. 8e-82;
Matches 455; Conservative 0; Mismatches 106; Indels 25; Gaps 1;

QY 2740 ttttggattttttgttttgagacagagcttgcctgttgcctgagccagccagagtcg 2799
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11794 TTTTGTGTTTGTGTTTGTGAGACAGAGTCTTTCTGTCAGCCAGGCTGGAGTGC 11735

QY 2800 agtggcagatctcagctcactgcacccctgccttcaggttcaactggttctctctgcc 2859
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11734 AGTGGCGCAATCTTGGCTCACTGCAACCTCTGCCTCTGGATTCAAGCAATATCTCTGCC 11675

QY 2860 tcagccctcagagtagttggattacaggtgccacacacacagcctggctaatatttga 2919
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11674 TCAGCCTCCTAGTAGCTGGGATTACAGGCGCCGCCACCGCTGGCTTAATTTTGCA 11615

QY 2920 tttttagtagacaggggtttcaccatgttgcagagactgctcaaaactcctgacctca 2979
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11614 TTTTGTAGAAACAGGGTTTCAACATGTTGCCAGGCTGCTCTCAACACCTGATCTCA 11555

QY 2980 agtgcacccacccctggcctccaaagtctgggattacagcagatgagccacgctgcc 3039
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11554 GGTATCCACCTGGCTTGGCTCCGCAAGTCTGGAATTACAGGCGTGAGCCACAGTGCC 11495

QY 3040 tggcctgtttttt- - - - - gaaatgaggtctggagtgca 3074
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11494 CAGCTTTTGTGTTTGTAGTAGGAGTCTGCTCCATCACCAGGCTGGAGTGCAG 11435

QY 3075 tggtcgatcatagttcactgcagctcaacctccagccagccagtgatctcctgcctc 3134
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11434 TGGTGAATCTCCACTCACTGCAACCTCTGCTCTCTGGTTCAAGTATTCTCTGCTC 11375

QY 3135 agcccttgagtagctggggtacagcgccacacacacatgcctggctagtttttaaaat 3194
||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11374 AGCTTCAACAAGTAGCTGGGACTACAGGCACATGTCACCATGCCAGCTAATTTTGTAA 11315

QY 3195 ttttgagagtaggttttactatgttgcaggtcaactctgaacctcctcgcttaag 3254
||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11314 TTTAGTAGAGCGGGTCTCAACACATTGGCCAGGCTGGTCTTGAACCTCTGACCTCAAG 11255

QY 3255 caaccctctgtctcagcctccacagtgctaggattacaagcgtg 3300
||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11254 TAATCCACCGCCTCAGGCTCCCAAGTGTGGGATTACAGGTGTG 11209

RESULT 3
AAI63970/C
ID AAI63970 standard; DNA; 10503 BP.
XX
AC AAI63970;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 342.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; vulnary; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ds.
OS Homo sapiens.
XX

PN WO200155308-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01309.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
PI

Qy	3072	cagtggtgcgatactagtctcaactgcagcgtccacgtccacgtccacggccccaaagtgatctctctgc	3131
Db	4786	cagtggcgatcatgctgcacctgcagctctactcttccaggaccagtgatctctcccac	4845
Qy	3132	ctcagcccttgtagtagctggggctacagcgccacacaccatgccctgctagtttttaa	3191
Db	4846	ctcagctctccaagtctgggactacagggccgaccaccatgcccgctaattttat	4905
Qy	3192	aatttttggagatgaggtttcacactatgtgttcagagctaatcttgaactcctcggctt	3251
Db	4906	atttttggtagaggaggtttcacacatgtgtccagcttgtgcaaacctctcgggctc	4965
Qy	3252	aagcaacccttggtctgcagctccacagtgctaggattacaagcgtg	3300
Db	4966	aggtgatccaccatcttggctcccaagtgtcaggattacaggcttg	5014
RESULT	5		
AAF97861/c			
ID	AAF97861	standard; DNA; 13864 BP.	
XX			
AC	AAF97861;		
XX			
DT	31-MAY-2001	(first entry)	
XX			
DE		Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO:75.	
XX			
KW		Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;	
KW		tumour suppressor; human lp36 homozygosity deletion domain; tumour;	
KW		diagnosis; ds.	
XX			
OS		Homo sapiens.	
XX			
PN	WO200116311-A1.		
XX			
PD	08-MAR-2001.		
XX			
PF	31-AUG-2000; 2000WO-JP05930.		
XX			
PR	31-AUG-1999; 99JP-0245962.		
PR	09-MAY-2000; 2000JP-0136266.		
XX			
PA	(HISM) HISAMITSU PHARM CO LTD.		
XX	(CHIB-) CHIBA PREPECTURE.		
PI	Nakagawara A;		
XX			
DR	WPI; 2001-226686/23.		
XX			
PT	Human lp36 homozygosity deletion domain from the 36-position of first		
PT	chromosome short arm in human neuroblastoma cell lines, applicable e.g.		
PT	in gene diagnosis of tumors as well as in developing anti-cancer drugs		
PT	.		
XX			
PS	Example 8; Page 143-149; 226pp; Japanese.		
XX			
CC	The present invention describes a homozygosity deletion domain		
CC	co-existing in the 36-position of the first chromosome short arm (lp36)		
CC	in human neuroblastoma. Also described are base sequences from the lp36		
CC	position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),		
CC	which are tumour suppressor genes in human neuroblastoma. The genes are		
CC	tumour suppressor genes, base sequence data of which are applicable as		
CC	tumour markers and reagents in studying mechanism of tumour body		
CC	formation, and gene diagnosis of tumours as well as in developing		
CC	anti-cancer drugs. AAF9787 to AAF97829 represent PCR primers used in		
CC	the exemplification of the present invention, and AAF97830 to AAF97874		
CC	represent sequences given in the exemplification of the present		
CC	invention.		
XX			
SQ	Sequence 13864 BP; 3536 A; 3244 C.3560 G; 3524 T; 0 other		

Query Match 10.1%; Score 333.2; DB 22; Length 1386A;
Best Local Similarity 75.5%; Pred. No. 5.7e-76;
Matches 453; Conservative 0; Mismatches 118; Indels 29; Gaps 2;

Qy 2728 agacctctcttttggtaatttttggtttgagacagagtcttgatcgttgtccc 2787
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2860 AGAAGTTTGAGTTTTAGATTTTTTGTGTTTTTGGACACGGAGTCTCGCACTGTCGCC 2801
Qy 2788 aggcacagtgtagtggcagcatctcagctcaactgcgaacctccgcttcagggttcaaact 2847
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2800 GGGCTGGAATGCAGTGGTGTGATCTCAGCTCAC TGCACTTCCGCCCTCCCGGTTCAAAGC 2741
Qy 2848 ggttcctcctcagcctccagagtagtggtgattaccagtgccccacacacagcctgg 2907
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2740 AATTTCCTGCCTCAGGCTCCAGA TAGTAGTGGGATTACAGGCGCCGACACACGCCACG 2681
Qy 2908 ctaattttgatatttagtagacagaggtttcaecatgttgccagactgggtctcaaa 2967
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2680 CTAATTTTTGATTTTTTAGTAGAGATGGGGTTTTACCATCTTGGCCAGGCTGGTCTCAA 2621
Qy 2968 ctctcgacctcaagtatccaccacacctgggccctcccaaatgctgggattcacaggcatg 3027
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2620 CTCCTGACCTC - GTGATCCGCCCGCCTTGCCCTCCCCAAAGTGCTGAGATTACCGCGCTG 2563
Qy 3028 agccacgctgctggcctggtttt-----ttgaatatga 3060
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2562 AGCCACGCTGCCCGCCCTTTTTTTCATGTCATTTGTGAGATGGAGTCTGGGTGCTCTCC 2503
Qy 3061 ggctgagtgtagtgatcgatcatgttcactgcagcctcaacctccagggcccaagt 3120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2502 AGGCTGGAGGCGAGTGGCGCATCTGGGCTCACTGCAACCTCCACTCCCAGGTTCAAAGC 2443
Qy 3121 gatcctcctgctcagcccttgagtagctgggctacagcgacacacacacacacgtcctgg 3180
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2442 GATTCTCCTGCCTCAGCCTCCCAAGTAGTGGAGTACAGGTGTGTGCCACCATGCCACG 2383
Qy 3181 ctagtttttaaaaattttgtgagatgaggtttcaactatgtgtccaggctaattttaa 3240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2382 CTAATGTGTTTTATTTTTAGTAGAGACAGGGTTTCAACCATGTTGGCCAGGCTGGTCTCGAA 2323
Qy 3241 ctctcggcttaagcaacctctggtctcagcctcccaagctgctagattacaagcgtg 3300
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2322 CTCCTAACCTCAAGTAGPCCACCCACCTCAGCCTCCCAAGTTCTGGGATTAGAGGTGTG 2263

RESULT 6
AAI62640/c
ID AAI62640 standard; DNA; 20444 BP.
XX XX
XX AAI62640;
XX AC
XX AC
DT 19-OCT-2001 (first entry)
XX DE
DE Human breast or ovarian antigen genomic DNA SEQ ID NO: 290.
XX KW
KW Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
XX ds.
XX OS Homo sapiens.
XX XX
XX WO20015324-A2.
XX PN
XX PD
PD 02-AUG-2001.
XX PF
PF 17-JAN-2001; 2001WO-US01344.
XX PR
PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 27-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488785/53.
New isolated nucleic acids and polypeptides, useful for diagnosing,
treating and/or preventing human diseases and disorders -
Disclosure; SEQ ID NO: 290; 520pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of ovarian and breast antigens. These are shown in
AA162467-AA162572 and AA42240-AA42345. The sequences can be used in the
diagnosis, prevention and treatment of breast and ovarian cancers, and
their metastases. The present sequence is a genomic sequence of the

RESULT	9
AAI62714	
ID	AAI62714 standard; DNA; 7588 BP.
XX	
XX	AAI62714;
XX	
XX	19-OCT-2001 (first entry)
XX	
DE	Human breast or ovarian antigen genomic DNA SEQ ID NO: 364.
DE	
KW	Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
KW	ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200155324-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01344.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488785/53.
New isolated nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders -
Disclosure; SEQ ID NO: 364; 520pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a number of ovarian and breast antigens. These are shown in AAI62467-AAI62572 and AAM42240-AAM42345. The sequences can be used in the diagnosis, prevention and treatment of breast and ovarian cancers, and their metastases. The present sequence is a genomic sequence of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO


```

FT exon /note= "potential 5' regulatory region"
FT /2001..2356
FT /*tag= b
FT /number= 1
FT intron 2357..3539
FT /*tag= c
FT exon 3540..3884
FT /*tag= d
FT /number= 2
FT intron 3885..12162
FT /*tag= e
FT exon 12163..12282
FT /*tag= f
FT intron 12283..15143
FT /*tag= g
FT exon 15144..15200
FT /*tag= h
FT intron 15201..15764
FT /*tag= i
FT exon 15765..15911
FT /*tag= j
FT intron 15912..19578
FT /*tag= k
FT exon 19579..19752
FT /*tag= l
FT intron 19753..19898
FT /*tag= m
FT exon 19899..19958
FT /*tag= n
FT intron 19959..20055
FT /*tag= o
FT exon 20056..20187
FT /*tag= p
FT intron 20188..20328
FT /*tag= q
FT exon 20329..20957
FT /*tag= r
FT intron 20958..21046
FT /*tag= s
FT exon 21047..21187
FT /*tag= t
FT variation 21188..21282
FT /number= 10
FT /replace(818,G)
FT /*tag= u
FT variation /standard_name= "single nucleotide polymorphism"
FT /note= "marker 17-2-297 (A31)"
FT /replace(1243,T)
FT /*tag= v
FT /frequency= 0.15
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-148 (A10)"
FT /replace(1374,A)
FT /*tag= w
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-256 (A32)"
FT /replace(1401,T)
FT /*tag= x

FT exon /frequency= 0.12
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-307 (A11)"
FT /delete(1535)
FT /*tag= y
FT variation /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-19-442 (A12)"
FT /replace(1788, A)
FT /*tag= z
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-20-187 (A13)"
FT /replace(2391, G)
FT /*tag= aa
FT variation /frequency= 0.24
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-1-308 (A14)"
FT /replace(3778,T)
FT /*tag= ab
FT variation /frequency= 0.29
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-3-324 (A15)"
FT /replace(4498,T)
FT /*tag= ac
FT variation /frequency= 0.22
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 99-14419-424 (A16)"
FT /replace(15007,A)
FT /*tag= ad
FT variation /frequency= 0.35
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-24-360 (A17)"
FT /replace(15233,A)
FT /*tag= ae
FT variation /frequency= 0.15
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-24-486 (A18)"
FT /replace(15826,T)
FT /*tag= af
FT variation /frequency= 0.01
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-6-187 (A19)"
FT /replace(19567,G)
FT /*tag= ag
FT variation /frequency= 0.35
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-148 (A20)"
FT /replace(19744,A)
FT /*tag= ah
FT variation /frequency= 0.14
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-325 (A21)"
FT /replace(19786,C)
FT /*tag= ai
FT variation /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-7-367 (A22)"
FT /replace(20158,G)
FT /*tag= aj
FT variation /frequency= 0.05
FT /standard_name= "single nucleotide polymorphism"
FT /note= "marker 9-9-246 (A23)"
FT /delete(20595..20597)
FT /*tag= ak
FT variation /frequency= 0.26
FT /note= "marker LSRX9-BM (17-1-240) (A24)"
FT /replace(21108,G)
FT /*tag= al
FT variation /standard_name= "single nucleotide polymorphism"
FT /note= "marker LSRX10-BM (A25)"
FT /replace(606,T)
FT /*tag= am
FT variation /standard_name= "single nucleotide polymorphism"
FT /note= "marker A'1"
FT /insert(5141,G)

```



```
FT primer_bind 1602..1621
FT /*tag= w
FT complement(1616..1635)
FT primer_bind /*tag= x
FT 1769..1787
FT misc_binding /*tag= y
FT allele replace(1788,C)
FT /*tag= z
FT misc_binding complement(1789..1807)
FT /*tag= aa
FT CDS 2001..21190
FT /*tag= ab
FT /product= "LSR"
FT exon /note= "this sequence contains introns"
FT 2001..2356
FT /*tag= ac
FT /number= 1
FT primer_bind 2036..2053
FT /*tag= ad
FT primer_bind 2062..2081
FT /*tag= ae
FT primer_bind complement(2074..2093)
FT /*tag= af
FT primer_bind 2084..2102
FT /*tag= ag
FT misc_binding 2372..2390
FT /*tag= ah
FT allele replace(2391,C)
FT /*tag= ai
FT misc_binding complement(2392..2410)
FT /*tag= aj
FT primer_bind complement(2563..2580)
FT /*tag= ak
FT primer_bind complement(2470..2489)
FT /*tag= al
FT primer_bind complement(2483..2500)
FT /*tag= am
FT primer_bind 3455..3474
FT /*tag= an
FT exon 3540..3884
FT /*tag= ao
FT misc_binding /number= 2
FT 3759..3777
FT primer_bind /*tag= ap
FT 3775..3792
FT allele /*tag= aq
FT replace(3778,T)
FT /*tag= ar
FT misc_binding complement(3779..3797)
FT /*tag= as
FT primer_bind complement(3882..3901)
FT /*tag= at
FT primer_bind complement(4336..4356)
FT /*tag= au
FT primer_bind 4444..4463
FT allele /*tag= av
FT replace(4498,G)
FT /*tag= aw
FT primer_bind complement(4902..4920)
FT /*tag= ax
FT misc_binding 4979..4997
FT /*tag= ay
FT misc_binding complement(4999..5017)
FT /*tag= az
FT misc_binding 5122..5140
FT /*tag= ba
FT misc_binding complement(5142..5160)
FT /*tag= bb
FT primer_bind 6638..6655
FT /*tag= bc
FT primer_bind complement(7072..7089)
FT /*tag= bd
FT misc_binding 7409..7427
```

```
FT misc_binding /*tag= be
FT complement(7429..7447)
FT /*tag= bf
FT 7995..8012
FT primer_bind /*tag= bg
FT misc_binding 8375..7393
FT /*tag= bh
FT allele replace(8394,C)
FT /*tag= bi
FT misc_binding complement(8395..9413)
FT /*tag= bj
FT primer_bind complement(8576..8593)
FT /*tag= bk
FT misc_binding 8685..8703
FT /*tag= bl
FT allele replace(8704,T)
FT /*tag= bm
FT misc_binding complement(8705..8723)
FT /*tag= bn
FT misc_binding 9009..9027
FT /*tag= bo
FT allele replace(9028,G)
FT /*tag= bp
FT misc_binding complement(9029..9047)
FT /*tag= bq
FT primer_bind 9622..9639
FT /*tag= br
FT misc_binding 9931..9949
FT /*tag= bs
FT misc_binding complement(9951..9969)
FT /*tag= bt
FT misc_binding 9958..9976
FT /*tag= bu
FT primer_bind 9984..9981
FT /*tag= bv
FT allele replace(9977,T)
FT /*tag= bw
FT misc_binding complement(9978..9996)
FT /*tag= bx
FT misc_binding 10002..10020
FT /*tag= by
FT allele replace(10021,G)
FT /*tag= bz
FT misc_binding complement(10022..10040)
FT /*tag= ca
FT primer_bind complement(10023..10040)
FT /*tag= cb
FT primer_bind 10492..10512
FT /*tag= cc
FT primer_bind complement(10546..10563)
FT /*tag= cd
FT misc_binding 11857..11877
FT /*tag= ce
FT primer_bind complement(10996..11015)
FT /*tag= cf
FT allele replace(11878,T)
FT /*tag= cg
FT misc_binding complement(11879..11897)
FT /*tag= ch
FT primer_bind 11972..11990
FT /*tag= ci
FT primer_bind 12005..12023
FT /*tag= cj
FT exon 12163..12282
FT /*tag= ck
FT /number= 3
FT primer_bind complement(12417..12436)
FT /*tag= cl
```

Query Match 9.88; Score 323.8; DB 22; Length 23187;
Best Local Similarity 75.58; Pred. No. 2e-73;
Matches 457; Conservative 0; Mismatches 112; Indels 36; Gaps 3;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:34:22 ; Search time 8779.43 Seconds
(without alignments)
4039.108 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300

Perfect score: 3300

Sequence: 1 tactaaaatacaaaattag.....gtgctaggattacaagcgtg 3300

Scoring table:

IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estfun:**

2: em_esthum:**

3: em_estin:**

4: em_estom:**

5: em_estpl:**

6: em_estba:**

7: em_estro:**

8: em_estov:**

9: em_htc:**

10: gb_estl:**

11: gb_est2:**

12: gb_htc:**

13: gb_gss:**

14: em_gss_fun:**

15: em_gss_hum:**

16: em_gss_inv:**

17: em_gss_pin:**

18: em_gss_pro:**

19: em_gss_rod:**

20: em_gss_vrt:**

21: em_gss_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	446.6	13.5	475	11	BF902291 IL2-MT017
C 2	371	11.2	371	11	BI021138 IL2-MT017
C 3	338.8	10.3	344	13	AQ069145 HS-2253_B
C 4	325.4	9.9	877	13	AQ739838 HS-5505_A
C 5	320	9.7	642	10	AL597806 DKF2P3130
C 6	317.4	9.6	679	11	BF346320 602018525
C 7	315.8	9.6	595	11	BF816106 MK2-C1012
C 8	309.8	9.4	603	10	AL138254 DKF2P547E
C 9	307	9.3	5528	13	AB012089 Homo sapi
C 10	305.6	9.3	603	10	AL138254 DKF2P547E
C 11	305.4	9.3	753	11	BG542878 602571053
C 12	304.4	9.2	963	11	BG677697 602625376

C 13	303.4	9.2	702	11	BF578427
C 14	303.4	9.2	856	13	AQ891549
C 15	303.2	9.2	660	13	AQ381918 RSC111-13
C 16	297.8	9.0	842	13	AQ748766 HS-5540_A
C 17	295.6	9.0	691	11	BG392240 602410376
C 18	295.6	9.0	1875	12	BC009270 Homo sapi
C 19	294.2	8.9	570	13	AZ537703 HSC 00199
C 20	294	8.9	658	13	AQ393450 CITB1-E1-
C 21	292.6	8.9	487	10	AI679002 tuc0901.x
C 22	292.2	8.9	588	13	AQ421340 RSC1-11-1
C 23	291.8	8.8	746	11	BG431326 60249916
C 24	291.8	8.8	5556	13	AQ839814 260L13-C5
C 25	289.2	8.8	605	11	BF872337 MRI-ET014
C 26	289	8.8	1345	10	AV762220 AV762220
C 27	288.6	8.7	767	10	AV700498 AV700498
C 28	288.2	8.7	684	11	BE908602 601496014
C 29	287.6	8.7	683	10	AV760485 AV760485
C 30	287	8.7	626	10	AV760760 AV760760
C 31	286.6	8.7	985	10	AL044339 DKF2P434A
C 32	286.2	8.7	834	11	BG530074 602558734
C 33	286.2	8.7	839	10	AV700988 AV700988
C 34	285.8	8.7	858	10	AU119532 AU119532
C 35	284.2	8.6	866	13	AQ894479 HS-3133-A
C 36	283.2	8.6	689	13	AQ318783 RSC111-92
C 37	282.4	8.6	814	13	AF102051 AF102051
C 38	281.4	8.5	523	10	A1054414 q176h03.x
C 39	281.4	8.5	611	10	AV763492 AV763492
C 40	281	8.5	644	11	BG426955 602493145
C 41	281	8.5	658	13	AQ357800 CITB1-E1-
C 42	280.6	8.5	626	10	AV760760 AV760760
C 43	280	8.5	792	11	BI225234 602949941
C 44	279.8	8.5	601	10	AV760723 AV760723
C 45	279.4	8.5	618	10	AV762702 AV762702

ALIGNMENTS

RESULT 1
BF902291/c

LOCUS BF902291 475 bp mRNA EST 18-JAN-2001
DEFINITION IL2-MT0179-121200-290-C04 MT0179 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF902291

VERSION BF902291.1 GI:12293750

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 475)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97. (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT 20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&tl2=IL2-MT0179-

121200-290-C04&tl3=2000-12-12&tl4=1)

Seq primer: puc 18 forward

sapiens genomic clone Plate=2253 Col=19 Row=N, DNA sequence.

ACCESSION AQ069145
 VERSION AQ069145.1 GI:3384344
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 344)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2253 row: N column: 19
 Class: BAC ends
 High quality sequence stop: 344.
 FEATURES
 source
 1..344
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2253 Col=19 Row=N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 67 a 105 c 82 g 88 t 2 others
 ORIGIN
 Query Match 10.3%; Score 338.8; DB 13; Length 344;
 Best Local Similarity 98.8%; Pred. No. 6.7e-31;
 Matches 340; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 2809 atctcagctcactgcaacctcgccctccagggttcaactggtctcctgctcagccctcc 2868
 Db 1 ATCTCAGCTCACTGCAACTCGGCTTCCAGGTTCAACTGGTCTCTCTGCTCAGCCTCC 60
 Oy 2869 agagtgtgtggattacaggtgccccaccacccagcctggctaaattttgtatttagta 2928
 Db 61 AGAGTAGTGGGATTACAGGTGCCACCACCACCGCTGGCTAAATTTTGTATTTTAGTA 120
 Oy 2929 gagacaggtttaccattgtggccagactggtctcaaacctcagctcagtgatcca 2988
 Db 121 GAGACAGGGTTTCACTGTTGGCCAGACTGGTCTCAACATCCCTGACCTCAAGTGATCCA 180
 Oy 2989 cccacctggcctcccaagtgcgtgggattacaggatcagccacccgctggcctggt 3048
 Db 181 CCCACTGNGCCTCCCAAGTCTGGGATTACAGCATGACCCACCGTGGCTGGT 240
 Oy 3049 ttttgaatgaggtctgagtgcaagtgtgcatcatagttcaactgcagcctcaacctc 3108
 Db 241 TTTTGAATGAGGTCTGGAGTGCAGTGGTGGATCATATGATGATGATGATGATGATG 340
 Oy 3109 ccaggcccaagtgatctcctcctcagcccccttgtagtctgg 3152
 Db 301 CCAGGCCCAAGTGATCTCTCTGCTCCTCAGCCNCTTGAGTAGCTGG 344
 RESULT 4
 AQ739838/c 877 bp DNA 16-JUL-1999
 LOCUS AQ739838
 DEFINITION HS_5505_A1_C09_47A RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=1081 Col=17 Row=E, DNA sequence.

ACCESSION AQ739838
 VERSION AQ739838.1 GI:5517360
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 877)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 1081 row: E column: 17
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 877.
 FEATURES
 source
 1..877
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1081 Col=17 Row=E"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
 BASE COUNT 251 a 181 c 242 g 203 t
 ORIGIN
 Query Match 9.9%; Score 325.4; DB 13; Length 877;
 Best Local Similarity 73.8%; Pred. No. 1.4e-29;
 Matches 442; Conservative 0; Mismatches 126; Indels 31; Gaps 1;
 Oy 2733 ttctctctttttgttatattttttgtttgtttgagacagagcttctgtgtccagagcc 2792
 Db 658 TCCTTTTCTTATTTATTTGTTTGTGTTTGTGAGACAGAGTCTCACTTTGATGCCAGGCT 599
 Oy 2793 agagtgcagtgccagcatctcagctcactgcaacctccgcttccaggttcaactggttc 2852
 Db 598 GGAGTACAGTGGTGAATCTCAGCTCACTGCAACCTTCGCCTCCAGGTTCAAGCAATTC 539
 Oy 2853 tctgtcctcagctccagagtagttgggattacaggttgcaccaccacgcctggctaat 2912
 Db 538 CCTTGCCCTCAGCTCCCAACTAGCTGGGACTACAGGTGCTTCCACCACCGCCAGCTAAT 479
 Oy 2913 tttgtatttttagtagagacaggtttccacatcttgccagactggtctcaactcct 2972
 Db 478 TTTTATATTTTATGATAGACAGAGGGTTTACCATGTTGGCCAGGCTGTGTTGAATCTCT 419
 Oy 2973 gacctcaagtgaatccaccacacctggcctccaaagtctgggattacaggtcagcca 3032
 Db 418 GACCTCAGCGCATCGCTAGCCTTGCCCTCCCAAAGTCTGGGATTATTAAGCGTGAGCCA 359


```

Qy 3206 tgaggttcaactatgtgtccaggctaaatctgaactcctggcttaagcaaacctctgg 3265
Db 512 CGGGTTTCCACATGTTGTGCGAGGCTGTGCTGAATCTTGACCTCAGGTATCATCCG 571
Qy 3266 tctcagctccacagtgcttagga 3289
Db 572 TCTCGGCTCCCAAGTGTGGGA 595

RESULT 8
AL138254/c
LOCUS AL138254 603 bp mRNA EST 01-MAR-2000
DEFINITION DKFP547E217_r1 547 (synonym: hfbr1) Homo sapiens cDNA clone
ACCESSION DKFP547E217 5', mRNA sequence.
VERSION AL138254
KEYWORDS AL138254.1 GI:6854935
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 603)
AUTHORS Ansoerge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Ansoerge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansoerge W
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFP547E217) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFP547E217"
/clone_lib="547 (synonym: hfbr1)"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="Xl-2blue"
/notes="vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT 112 a 166 c 156 g 168 t 1 others
ORIGIN
Query Match 9.4%; Score 309.8; DB 10; Length 603;
Best Local Similarity 75.9%; Pred. No. 1.2e-27;
Matches 431; Conservative 0; Mismatches 113; Indels 24; Gaps 3;

Qy 953 ctttggaggccgagcgagtggaattacctgaggtcaggagttcaagaccagcctggcca 1012
Db 603 CTTTGGGAGGACGAGCGGGCAATCACTGAGTCAGGAGGAGTTCAGACCCATAGCCA 544
Qy 1013 acatggagaaacccagtcctctaatataaaaaaaattagccaggtgtgtggcca 1072
Db 543 ACACGGTGAACCCCATCTACT-AAAATACAAAAATTAGCTGGGTGTGGTGGCT 488
Qy 1073 catgcttgtaatccagctactcaggagactaacgcaagagaattgttgaaaccagagag 1132
Db 487 CACACCTGTAGTCCCGACTACTCGAGAGGCTGAGCGAGGAGAAATGCTTGAGCCCTGGAG 428
Qy 1133 gcagaggttggtgagccgagatcgcgccatttcactccagcctgggaaacaaagagcga 1192
Db 427 GTGAGGTTGAGTGAGTGAATCGGCCACTCGCCCTCCAGCCTCGGCCACAGACAA 368
Qy 1193 gactccatctcaaaaaaa- - - - -gtgggagcgagagcgagag 1236

```

```

Db 367 GACCTGTCTCAAAACAACACACAGTCCCAGGACTTTGGGAGCCGAGGCGAGCA 308
Qy 1237 gatcactagagccagtagtatttgagaccatcctctgggcaacatagcaggaacctgtctgta 1296
Db 307 GATGGCTTGAGCCCGAGGAGTTTGAGACCAAGCTGGGCAACATGTGTAACCCCTGTCCCTG 248
Qy 1297 caaaaaataaaaaaattaaacgggcatggtgggcacacacccgtagtcccagctact 1356
Db 247 CAATA- - - - -CATAAAAAATTTGGCGGCGGTGGTGGCCACACCTGTAGTCCGCGCTACT 192
Qy 1357 ccagaggtgagggcagggagctcgtgagccaggaatttgaggctgcagtgaactgtg 1416
Db 191 TTGAGAGCTGAGGGGGAGGATAGCTTCAACCCGGAGGCGAGGCTCAGTGAGCTGTG 132
Qy 1417 atcccaccactgcacttaagcctggaatacaaaagacacccctgtctcaataacaatag 1476
Db 131 ATCGCACCACTGCATCCAGCTGGGTGACAGAGGAGACCCCGTCTTNAACCAACCA 72
Qy 1477 caataataataaaaaaattaaatgca 1504
Db 71 AACCAAGCAAAACAACAAAAACCCA 44

RESULT 9
AB012089
LOCUS AB012089 5528 bp DNA GSS 21-MAY-1998
DEFINITION Homo sapiens gene, genomic survey sequence.
ACCESSION AB012089
VERSION AB012089.1 GI:3149956
KEYWORDS GSS.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5528)
AUTHORS Kurahashi,H.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hiroki
Kurahashi, Biomedical Research Center, Osaka University Medical
School, Division of Clinical Genetics, Department of Medical
Genetics; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:kura@cigene.med.osaka-u.ac.jp, Tel:81-6-879-3381,
Fax:81-6-879-3389)
REFERENCE 2 (sites)
AUTHORS Kurahashi,H., Sakamoto,M., Ono,J., Honda,A., Okada,S. and
Nakamura,Y.
TITLE Molecular cloning of the chromosomal breakpoint in the LIS1 gene of
a patient with isolated lissencephaly and balanced t(8;17)
JOURNAL Hum. Genet. (1998) In press
FEATURES
Location/Qualifiers
source
1..5528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13.3"
BASE COUNT 1468 a 1190 c 1182 g 1688 t
ORIGIN
Query Match 9.3%; Score 307; DB 13; Length 5528;
Best Local Similarity 74.2%; Pred. No. 6.8e-28;
Matches 443; Conservative 0; Mismatches 120; Indels 34; Gaps 3;

Qy 2733 tttctttttttgtattttttttttgtttgagacagagttgatctgtgtgccagggc 2792
Db 2521 TTTCTCCAGATTCCTTTTTTTTTTTTTTTTCTTGACACGAGTCTCTCTGCTCGCTAGACT 2580
Qy 2793 agagtgcagtgccagctcagctcactgcaacctccgctccaggttccaactggttc 2852
Db 2581 GGATGAGTGAGCGACAGACTCTGGCTTACTGCGCTCCACCTCCAGGTTGGA- - - - -GTGC 2637
Qy 2853 tctgctcagcctccagagtagttgggattacaggttgccaccaccagcgtggtta 2912

```


High quality sequence stop: 693.
 Location/Qualifiers
 1. 753
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4695636"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (TI phage-resistant)"
 /notes="Organ: lung; Vector: pDNr-LIB (Clontech); Site:1: SfiI (ggccattggcc); Site:2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
 BASE COUNT 211 a 188 c 192 g 161 t 1 others
 ORIGIN

Query Match	9.3%	Score 305.4;	DB 11;	Length 753;
Best Local Similarity	74.2%;	Pred. No. 3.4e-27;		
Matches	422;	Conservative	0;	Mismatches 126;
			Indels	21;
			Gaps	2;
QY 2750	ttttttgttttgagacagagtccttgatctgtgtgccaggccagagtgagtgagcaga	2809		
Db				
QY 593	TTCTTTTGTGAGACAGAGTCTAGCTTTGTCGTCCAAGCTGGAGTATGTTGGCGGTGA	534		
Db				
QY 2810	tctcagctcactgcgaactccgccttccaggttcaactggttcttcctgcctcagcctcca	2869		
Db				
QY 533	TCTCAGCTCACTGCAACTCGGCC-TCTGGGTTTCAACGCAATCTCATGCCTCGCCCTCCA	475		
Db				
QY 2870	gagtagttgggattacagggtgcacacacacagcctggctaaatttttgatttttagtag	2929		
Db				
QY 474	GGGTAGCTGGGATTACAGGCGCGTACCAACACACACCACTAAATTTTGTATTTTGGTAC	415		
Db				
QY 2930	agacagggtttccacctgtgtggccagactggctcacaactcctcgaactcaagtgtaccac	2989		
Db				
QY 414	ACAGGGGTTTCAACCATGTTGCCAGGCTGGTCTCGAATCTCTGGGCTGTAGTGATCCAC	355		
Db				
QY 2990	ccacctgggacctcccaagtgcgtgggattacaggcatagccaccgtgcctggcctgttt	3049		
Db				
QY 354	CCACCTTTGGGCTCCCAAGTCTGGGATTACAGCGGTGAACCAACATGCCAAGCTTTT	295		
Db				
QY 3050	ttttgaaatgag-----gtctggagtgaagtgcgatacatagt	3089		
Db				
QY 294	TATTTTGAGATGGAGTCTTGCTGTGTCGCAGGTTGGAGTGCAGAGCAATCTCGGC	235		
Db				
QY 3090	tcactgcagctcaaacctccacagcccaagtgcctctcctgcctcagcccttgagtagc	3149		
Db				
QY 234	TCACTGCAACTTCGGCTTCCGGGTTCAAGGGAGTCTCCTGCCTCAGCTTCTTGAGTAGC	175		
Db				
QY 3150	tggggctacaggcgacacacacacatgcctggctagtttttaaaattttgtggagatgag	3209		
Db				
QY 174	TGGGACTACAGGAGAGATGCCACACAGCCCAAGTATAGTTTTTGTATCTTGGTAGAGCAGG	115		
Db				
QY 3210	gttcacatgttgtcaggactaatcttgaaactcctcgccttaagcaacctctggctc	3269		
Db				
QY 114	GTTTACCATATTTGGCCAGGCTGGTCTTTGAACCTCTGGCCTCAAGTGTATCCCTTGCTC	55		
Db				
QY 3270	agcctccacatgctgtaggattacaagcg	3298		
Db				
QY 54	GGCCTCCCAAAGTGCTGGGATTACAGGAG	26		
Db				

RESULT 12
 BG677697
 LOCUS 963 bp mRNA EST 01-MAY-2001
 DEFINITION 602625376F1 NC1_CGAP_skn4 Homo sapiens cDNA clone IMAGE:4750146 5',
 mRNA sequence.
 ACCESSION BG677697

```

Qy 3192 aatttttgtagagagg-tttcaactatgttgcagggcctaatttgaaactcttcggct 3250
Db 587 -GTTTGTAGTAGAGGGGATTTTCCCATGTACACAGGCTGGTCTCGAACTCTTGAGCT 645

Qy 3251 taagcaacctctggtctcagcctccacacgtctagattacaagcgtg 3300
Db 646 CAAGCAATCTCCCATCTTGGCTCTCAAGAGCGTGGGATTATAGGCATG 695

RESULT 13
LOCUS BF678427 702 bp. mRNA EST 21-DEC-2000
DEFINITION 6020586ZF1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4250246 5',
mRNA sequence.
ACCESSION BF678427
VERSION BF678427.1 GI:11952322
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHCMI071 row: k column: 15
High quality sequence stop: 562.
FEATURES
    source
        1..702
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4250246"
            /lab_host="NIH_MGC_83"
            /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
            Site_1: SfiI (ggcccgtcgcc); Site_2: SfiI (ggccattatggc
            ); 5' and 3' adaptors were used in cloning as follows: 5'
            adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
            sequence: 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)BN-3'
            (where B = A, C, or G and N = A, C, G, or T). Average
            insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
            contained inserts by PCR. This library was enriched for
            full-length clones and was constructed by Clontech
            Laboratories (Palo Alto, CA)."
BASE COUNT 223 a 170 c 210 g 99 t
ORIGIN

Query Match 9.28; Score 303.4; DB 11; Length 702;
Best Local Similarity 76.3%; Pred. No. 6e-27;
Matches 442; Conservative 0; Mismatches 121; Indels 16; Gaps 5;

Qy 2734 tctctttttgtattttttgttttgagacagatcttgatct-gttgccagggcc 2792
Db 595 TTTTCTTTTCTTCTGTTCTTTCTTTTGTAGACAGAGCTCTACCCCTGGTGGCCAGTCT 526

Qy 2793 anagtcagtgagcagatctcagctcactgcgaacctccgctcttcaggttcaactgggtc 2852
Db 525 GGAGTGTCTATGGCGCAATCTCGGCTCACTGCAGCCTCCGCTTCCGGGTTCAAATGATTC 466

Qy 2853 tctgcctcagcctccagcagtagtggattacaggtgccccacacacgcgctgctaata 2912
Db 465 TCCTGCTTCAGCCTCCCGAGTAGCGGGATTACAGGTGCCCGCCACCATGCCCCAGCTAAT 406

```

```

Qy 2913 ttttgatttttagtagagacagaggttttccacctgttggccagactgggtctcacaactct 2972
Db 405 TTTTGTATTTTGTAGTAGAGACAGAGGTTTTCACCATGTTTACCAGGCTGGCTTCGAACCTCT 346

Qy 2973 gacctcaagttaatcacccacctgggctccaaaagtctgggattaccagcagcagcca 3032
Db 345 GACCTC--GTGATCCGCCCGCTCGGCTCCCAAAGTCTGGGATTACAGGTGTGAACCA 288

Qy 3033 ccgtgcctggcctgtgttttttgaatgaggtctgaggt-----gcagtgggtgctg 3081
Db 287 CTGTGCCCGCGCTTTTGTGTTTGTGTAGTCTGCTCTGTCGCGAGTGGTGG 228

Qy 3082 atcatgttctactcagcctcaacctccagggcccaagtatctctctcgtcctcagccct 3141
Db 227 ATCTCGGCTCACTCAACCTCTGCTCTTGGATTCAAGCGATTCTCCCTGCGCTCAGCCCTT 168

Qy 3142 tgaatagctggggtctacagggcgcaacacacacatgcctcgtgtagtttttaaaattttgtg 3201
Db 167 CCAGTAGCTAGGACTACATGCTGCATCTCCACGCTCGCTAAATTTTGTGTTTGTGTTA 108

Qy 3202 gagatgaggtttcactatgttctcaggtctaactctgaacctcctcgcttaagcaacct 3261
Db 107 GAGACG--GGTTTCAACATGTTG--CCAGGCTGCTCTCGAACTCTCGGCTCAAGTCATCTG 50

Qy 3262 ctgctctcagcctccacagtcgttaggattacaagcgtg 3300
Db 49 CCCATCTTGGCTTCTCTAAAGTCTGGGATTATAGGTGTG 11

RESULT 14
LOCUS AQ891549/c 856 bp DNA GSS 10-NOV-1999
DEFINITION HS_3143_AL_C12_T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3143 Col=23 Row=E, DNA sequence.
ACCESSION AQ891549
VERSION AQ891549.1 GI:6347739
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3143 row: E column: 23
Seq primer: T7
Class: BAC ends
High quality sequence stop: 856.
FEATURES
    source
        1..856
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=3143 Col=23 Row=E"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelOBAcl1; BAC Clones in
            E-Coli DH10B"
BASE COUNT 239 a 179 c 233 g 200 t
ORIGIN

```

2009-01-14 11:23:01

THIS PAGE BLANK (USPTO)
THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 15:53:12 ; Search time 14958.6 Seconds
(without alignments)
3639.417 Million cell updates/sec

Title: US-09-820-005-3_COPY_1_3300
Perfect score: 3300
Sequence: 1 tactaaatacaaaattag.....gtgctagattacaagcgtg 3300

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htgo_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description	
		Match	Length	DB			
1	3288.8	99.7	124526	9	AC005080	AC005080 Homo sapi	
2	3267.6	99.0	169604	9	AC083884	AC083884 Homo sapi	
3	3267.6	99.0	230552	9	AC005098	AC005098 Homo sapi	
4	3265.6	99.0	131359	9	AC004883	AC004883 Homo sapi	
5	3265.6	99.0	275197	9	AC004166	AC004166 Homo sapi	
6	3172	96.1	17302	9	AF184614	AF184614 Homo sapi	
c	7	2179.8	66.1	184558	2	AC092405	AC092405 Papio cyn
8	2106.8	63.8	8119	9	HS47L14S01	U61224 Homo sapien	
9	2105.8	63.8	8131	9	HS47PHOX01	U57833 Homo sapien	
10	1978	59.9	3074	9	AF003533	AF003533 Homo sapi	
11	1943.4	58.9	1965	9	AF184613	AF184613 Homo sapi	
12	1552.2	47.0	1569	9	HS47P41S01	U61241 Homo sapien	
13	1263.6	38.3	7258	9	HS47P43S01	U60970 Homo sapien	
14	1260.4	38.2	7258	9	HS47P40S01	U61238 Homo sapien	
15	1251.6	37.9	3594	9	HSNCF1S1	U69639 Homo sapien	
c	16	1090.4	33.0	149830	2	AC027219	AC027219 Homo sapi
17	661.2	20.0	57245	2	AC068263	AC068263 Homo sapi	
c	18	632.2	19.2	57245	2	AC068263	AC068263 Homo sapi
19	589.2	17.9	101509	2	AC027353	AC027353 Homo sapi	
20	481.4	14.6	524	9	HSU33006	U33006 Human p47-p	
21	462.8	14.0	486	9	HS39NCF1	U72356 Homo sapien	
c	22	437.4	13.3	198410	2	AP000831	AP000831 Homo sapi
23	416.6	12.6	186925	2	AC087503	AC087503 Homo sapi	
24	413.2	12.5	140161	2	AL390725	AL390725 Homo sapi	
25	410.4	12.4	115079	2	AC088105	AC088105 Homo sapi	
c	26	407.8	12.4	165341	2	AC092037	AC092037 Homo sapi
27	406	12.3	67674	2	AC090641	AC090641 Homo sapi	
c	28	406	12.3	108369	17	AF124523	AF124523 Homo sapi
29	406	12.3	344150	9	AF235103	AF235103 Homo sapi	
c	30	401.2	12.2	218445	2	AC083959	AC083959 Homo sapi
c	31	400	12.1	177990	2	AC016385	AC016385 Homo sapi
32	394.4	12.0	177990	2	AC016385	AC016385 Homo sapi	
33	394	11.9	194938	2	AC073841	AC073841 Homo sapi	
c	34	391.8	11.9	106928	9	AC005049	AC005049 Homo sapi
c	35	391.6	11.9	155046	2	AC069408	AC069408 Homo sapi
36	390.4	11.8	148822	9	AC020904	AC020904 Homo sapi	
37	390.2	11.8	42156	2	AC011542	AC011542 Homo sapi	
38	389.6	11.8	166181	9	HS37E16	283844 Human DNA s	
39	387	11.7	273403	2	AC011498	AC011498 Homo sapi	
40	386.6	11.7	11954	9	AL133275	AL133275 Human DNA	
41	385.8	11.7	41572	9	AC006132	AC006132 Homo sapi	
42	383.4	11.6	64681	2	AC023215	AC023215 Homo sapi	
43	382.8	11.6	198410	2	AP000831	AP000831 Homo sapi	
c	44	380.6	11.5	190225	2	AC011191	AC011191 Homo sapi
c	45	379.6	11.5	120955	9	HUAC002310	AC002310 Human Chr

ALIGNMENTS

RESULT 1
AC005080
LOCUS AC005080 124526 bp DNA PRI 02-OCT-2000
DEFINITION Homo sapiens BAC clone CTA-269P13 from 7q11.2, complete sequence.
ACCESSION AC005080
VERSION AC005080.2 GI:7770715
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 124526)
MEDLINE Sulston, J.E. and Waterston, R.
AUTHORS Toward a complete human genome sequence
TITLE Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL 99063792
REFERENCE 2 (bases 1 to 124526)
AUTHORS Scott, K., Layman, D., Kalicki, J. and Harmon, G.
TITLE The sequence of Homo sapiens BAC clone CTA-269P13
JOURNAL Unpublished

[illegible]

source	1. .169604 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="RP11-81337"	BASE COUNT 44465 a 38383 c 38794 g 47962 t	ORIGIN
Query Match	99.0%; Score 3267.6; DB 9; Length 169604;		
Best Local Similarity	99.8%; Pred. No. 0;		
Matches 3293; Conservative	0; Mismatches 4; Indels 3; Gaps 2;		
Qy 1	tactaaaaatacaaaatagccaggctgggtggcgacacctgtaataccagctacttgg 60		
Db 94621	TACTAAAAATACAAATATAGCCAGCGCTGGTGGCGCACACCTGTAAATCCCGACTACTTGG 94680		
Qy 61	gaactgaggcagagagaatcgcttgaacctggaaggcagagaggttgcagtgagccgagatt 120		
Db 94681	GAAGCTGAGGCGAGGAGAAATCGCTTGAACCTGGAAAGGAGAGGTTGCCAGTGAGCCGAGATT 94740		
Qy 121	gtgccactgcactccagcctggcgcaacaagagcgaaacttcgcttcaaaacaaataaatta 180		
Db 94741	GTGCCACTGCACCTCCAGCTTGGGCAACAAGACGCAAACTTCGCTTCAAAACAATAAATTA 94800		
Qy 181	acgccagcatgtcttggcttcaatctgccagacctcaacccctcaccccccaggagatcag 240		
Db 94801	ACGCCAGCATGTCTGGCTTTCATCTGCCAGACCTCAACCCCTCACCCCGAGGATCAG 94860		
Qy 241	gtccggaccatgaactgacctggaactcaggcaagggtgagttgagtcagccctggcctg 300		
Db 94861	GTCCGGACCATGAGCTGACCTGACCTCAGGCAAGGGTGAAGTTGGTGCAGGCCCTGGCCTG 94920		
Qy 301	ctgggagccacagctgcagcaggtgcctgggctgggctgagggcccgccactcatgaactcat 360		
Db 94921	CTGGAGGCACAGGCTGCAGCAGGCTGCCTGGGGCTGAGGGCCGCCACTCATGAACTCAT 94980		
Qy 361	gaccttgaatgagctccaaagctctgggctcccgctccaggtctagggggagtgggagagag 420		
Db 94981	GACCTTGAATGAGCTCCAAAAGCTCTGGGCTCCAGGCTCTAGGGGGAGTGGGAGAGAG 95040		
Qy 421	aggctcagcctgtccctgggcatgtgcctccctccctcaccctcttctcccaaatccctc 480		
Db 95041	AGGCTTCAGCCTGTCCCTGGGCACTGTGCCGCCCTCCCTCACCTCTTGTCCCAATCCCTC 95100		
Qy 481	tcttggcaaaegtacagctcttaataatacactctgagaaaaactgagtcagccctaaagaa 540		
Db 95101	TCCTGGCAAAAGCTCACAGCTTTATATATACCTCTCGAGAAAAGTACAGCCCTTAAGAA 95160		
Qy 541	caattcaatgaaccatttgccttacttgagattggaactcaagtctcactcaaaagtctgt 600		
Db 95161	CAATTCAATGAACCATTTTGCTTACTTGAGGATTGGAACCTCAAGTCTCACATCAAGTCTGT 95220		
Qy 601	gcaatttgcctcagctgtcactggccctcactccacacacacccaaaggatgagcatcta 660		
Db 95221	GCCATTTTCGTCCCAGCTGTCACTTGGCCCTCATCCACACACACCAAGGATGAGCATCTA 95280		
Qy 661	acgctgtcatgcacactcccatgcccgcttcattcactcaattcattcattcattcactc 720		
Db 95281	ACGCTGTCAATGCACACTCCCATGCCCGGCTTCAATTCATCTATTCAATTCATTCACTAC 95340		
Qy 721	attcattgactcattcattcattcactcactcattcattcactcactcagtgaaattgtgcagtc 780		
Db 95341	ATTCATTGACTCATTTTCATTCATTCACCTCACTCATTTTCATTCAGTGAATGTTGCAGTC 95400		
Qy 781	acgatcaaaatatttatggcctctgtgtgccaggcactagatgaggggctgggggctaga 840		
Db 95401	ACGATCCAAATATTTATGGCTCTGTGTGCCAGGCACTAGATGAGGGGCTGGGGCTAGA 95460		
Qy 841	gccccgataaccgggtcatgcccctagcttctctgggacacacaaattgtgtaaggaggaga 900		
Db 95461	GCCCCGTGATAAACCGGGTCATGCCCTAGCTTTCCCTGGGACACACATTTGTGTTAAGGGAGA 95520		
Qy 901	ctaaaaaaattaaagtcaggccagcggtggctcatgctcgtgaatcccgacctttggga 960		
Db 95521	CTAAAAAATTAAGTCAGGCAGGCACGCTGGCTCATCCCTGAATCCCGACACTTTTGGGA 95580		
Qy 961	ggcggaggcgagtgaaattacactgaggttcaggaggttcaagaccagcctgccaacaatgag 1020		
Db 95581	GGCGAGGCGAGTGAATTTACCTGAGGTGAGGTTCAGAGCAGCCTGGCCCAACATGAG 95640		
Qy 1021	aaacccagctctctaattaaaaaataaaaaaaattagccaggtgtggtggcacaatgcctg 1080		
Db 95641	AAACCCAGCTCTCTAATT--AAAAAATAAAAAAATTAACCAAGGTGTGGTGGCACAATGCCCTG 95698		
Qy 1081	taatccagcttactcagagactaacgcaagagaattgtcttgaaccccgaggcgagaggt 1140		
Db 95699	TAATCCCAAGCTTACTCAGGAGACTAACGCAAGAGAATTCCTTGAACCCAGGAGCGAGAGT 95758		
Qy 1141	tgcggtgagccgagatcgccattgcactccagcctgggaaaaacaagagcgagacatcat 1200		
Db 95759	TGCGGTGAGCCGAGATCGCCATTTGCACCTCCAGCTGGGAAACAAGAGCGAGACTCCAT 95818		
Qy 1201	ctcaaaaaaataaaagtgggagggcagagggcagaggtacactagagggccagtagtttga 1260		
Db 95819	CTC-AAAAAATAAAAGTGGGAGGCGAGGCGAGGAGGATCACTAGAGGCCAGTAGTTTGA 95877		
Qy 1261	gaccatcctgggcaacatagcagaccctgtctgtacaaaaaattaaaaaaatttaac 1320		
Db 95878	GACCATCTGGGCAACATAGCAGGACCCCTGTCTGTACAAAAAATTAATAAAAAATTTAAC 95937		
Qy 1321	cggcattgtggcacacacacccgttagtcccgactactccagggctgagggagagatcag 1380		
Db 95938	CGGCATGTGGTGGCACACACCCGTAGTCCAGCTAGTCCAGAGGCTGAGGCGCAGGAGATCG 95997		
Qy 1381	ctggagcccgaggtgtggaggtgcagtgaaactgtatccaccactcactcactaaagcctg 1440		
Db 95998	CTGGAGCCACAGGAGTTGGAGGCTGCAGTGAAGTGTGATCCCACTCCCACTGCATTAAGCCTG 96057		
Qy 1441	gatacaaaagcaagaccctgtctcaaaataacaaatagcaataataataaagaaaaattaaa 1500		
Db 96058	GATAACAAAGCAAGACCCCTGTCTCAAAATAACAATAGCAATAATAATAAAGAAAAATTA 96117		
Qy 1501	tgaatttgcagatgcatacagtgataagtgctctgcagaaaaagagagcgaggaagctg 1560		
Db 96118	TGCAATTTGCCATGCATCAGTGATAAGTGTCTGCAGAAAAAGGAGCGCAGGAAGGCTG 96177		
Qy 1561	agaaagtgataggttctgtctgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaat 1620		
Db 96178	ACAAGGTATCAGGTTTCTATGCAATGTGAAGTTATCAAGGAAGGCTTCTCGGAAGAG 96237		
Qy 1621	tgacatttgacagagaaatggagagaggttatggagggaagatgggtgaaatggggggaac 1680		
Db 96238	TGACATTTGACAGAGAAATGGAGGAGAGTTATGGAGGGAAGATGGTGAATGGGGGGAAC 96297		
Qy 1681	atggtcaagacaggaataatagttcaaggggggaaagatgggtcaaggggagcgagcaaatg 1740		
Db 96298	ATGTTCAAGACACAGGAATATGTTCAAGGGGGGAAAGATGGTCAAGGGGACGCGACAAATG 96357		
Qy 1741	caagggcctcagcgagcagcagcttgaattcaccccccaaacccgtgggcccgtgcag 1800		
Db 96358	CAAGGGCCCTGAGCGAGGAGCAGCTTGATTCACCCCCAAAACCCGTGGGGCCCTGTGCAGG 96417		
Qy 1801	cgaagggaagacaaagtataaaccccttttccctgttccctgcaggtgtgtgtgaaacatgag 1860		
Db 96418	CGACGGGAAGCACAAAGTGTAAACCCCTTTTCTCTGTCCTCGCAGGTGTCTGTGAACATGAG 96477		
Qy 1861	tctgccccatgtttcacacctgcagcctgaagagtcccccgaaacactgaagaagaagca 1920		
Db 96478	TCTGCCATGTTTACACCTGCAAGCCTTGAAGAGTCCCGACAAACTGAAGAAGCA 96537		
Qy 1921	agcccttctgacccctccctgccttcccgaccgcgacaaagcagcttccctctt 1980		
Db 96538	AGCCCTTCTGTACCCCTCCCTGCCCCCTGTCCCGACCCGACAAAAGGACTTCTCTCTT 96597		
Qy 1981	ccagtgcatataaagcgcgagccttgaaagtcgacggagagcactggagggccacccagtcag 2040		

```
Db 96598 CCACTGCACTTTAAGGCGCAGCGCTTGGAAAGTGCCAGGGAGCACTGAGGAGCCACCCAGTCAATG 96657
Qy 2041 ggggacaccttcacccgacacatcgccctgctgagcctttagaagccttcgtaccacgc 2100
Db 96658 GGGACACCTTCATCCGTCACATCGCCCTGCTGGGCTTTGAGAAAGCGCTTCGTACCCAGC 96717
Qy 2101 cagcactatgtagtagctggtgagggcattcccgctggggggaatacgggagggacagca 2160
Db 96718 CAGCACATATGTAGTAGTGTGTGAGGGCATCCCGTGGGGGATACGGGAGGCACAGCA 96777
Qy 2161 cggcacaccttgacgtccacgggccaaccagctccagtgaggaactaacggggcagggtct 2220
Db 96778 CGGCACACCTTTCAGTCCAGGGGCAACACAGCTCCAGTGGAGTAACAGGGGCGAGGCTCT 96837
Qy 2221 tggcaccttgctccctgctgctttagccttgatctacccctctgattccctgggaagacag 2280
Db 96838 TGGGCACTGTGCTGCTGCTTTGAGCCTGGATCTACCCCTCTGATCCCTGGGAGACAG 96897
Qy 2281 ttcccttgaccccgccctgggcccacgccttttactgtcccccgcctgtgtcccccagcag 2340
Db 96898 TTCCCTTGGACCCCGCTGGGCCCCAGCCCTTTACTGTCCCCGCTGTGTCCCCAGCCAG 96957
Qy 2341 gccctcagccttagcaggagctctcttctgtctccctgctgcatggccagcagccacgc 2400
Db 96958 GCCCTCAGCCTTAGCCAGGAGTCTCTTTCTGCTCCCTGCTGCAATGGCAGGACGCCAGC 97017
Qy 2401 gctctcagctccgagcccaactctccaggaaagccttccctgaactagccagctatca 2460
Db 97018 GCTCTCTAGCTCCAGGCCCACTCTCCAGGAAGCCTTCCTGACTAGCCAGCTATCA 97077
Qy 2461 gagagtgccctcccaagagggagccttggaaactaaagctctctctctcccccagctgcc 2520
Db 97078 GAGAGTGGCCCTCCCAAGAGGAGCGCTGGAAACTAAAGCTCTCTCTCTCCCGAGCTGCC 97137
Qy 2521 tgtagtgctcagtagagcttctctctccagtgagtgagtgagcagcagggcccaata 2580
Db 97138 TGTAGTCTCAGTTAGAGTCTTATCTCTCCAGTAGGCTGACACCATGACAGGGGCCAATA 97197
Qy 2581 gagtccctcccatctgtcccccagggagcgtggacaaatgcctgctcagacacacaagtcca 2640
Db 97198 GAGTCTCTCCCATCTGTCCCCAAGAGGCTGGACAAATGCTGCTCAGACACACAAGTCCA 97257
Qy 2641 ctgggtccctcctaattcccataggaagccagggaggaactacattaggaattgaagctt 2700
Db 97258 CTGGGTCCCTTAATCCCATAGGAAGGCCAGGAGGAACTACATTTAGGAATTTGAAGCTT 97317
Qy 2701 gtaggaacatttagctctatgtcccaagaccttctctttttttttttttttttttttgtt 2760
Db 97318 GTATGGAACTTTAGTCTATGTGCCAAGACCTTTCTCTTTTGTATTTTGTGTTTT 97377
Qy 2761 ttgagacagagctcttgatctgttggccagggccagagtgacagtgacagctcagctcac 2820
Db 97378 TTGAGACAGAGCTCTGTATCTGTTCGCCAGGCCAGAGTGCAGTGGCAGCATCTCAGCTCAC 97437
Qy 2821 tgcacacctccgcttccaggttccactggttctcctcctcagcctccagctcagtagttggg 2880
Db 97438 TGCAACCTCCGCTTCCAGGTTCAACTGGTTCTCTGCTCAGCTCCAGAGTAGTGGG 97497
Qy 2881 attacaggtgcccaccacacagcctgctaatttttatttttttttttttttttttttttt 2940
Db 97498 ATTACAGGTGCCCAACACACCGCTGGCTTAATTTTGTATTTTGTATGAGACAGGGGTTT 97557
Qy 2941 caccatgttgccagagctggtctcaaacctcctgacctcaagtgcacatgcacacccacctgggcc 3000
Db 97558 CACCATGTTGCCAGAGCTGCTCAAACTCTGACCTCAAGTGATCCACCCACCTCGGGCC 97617
Qy 3001 tcccaagtgctgggattacaggcatgagccaccgtgctggcctgtttttttttttttttttttt 3060
Db 97618 TCCCAAAAGTGTGGGATTACAGGATGAGCCACCGCTGCTGGCTGTGTTTTTTTGAATGA 97677
Qy 3061 ggtctgagtgagtggtgcatcatagtttcaactgacgtccacccctccagccagcccaagt 3120
Db 97678 GGTCTGGAGTGCAGTGGTGGCATCATAGTTTCTACTGACGCTTCGACCTCCCGAGGCCCAAGT 97737
```

```
Qy 3121 gatcctcctgctcagcccttgagtagctggggtacagggccacacccaccatgctcgg 3180
Db 97738 GATCCTCTGCTCAGCCCTTGAGTACCTGGGCTACAGGCGCACACCACCATGCTCTGG 97797
Qy 3181 ctagtttttaaatttttggagatgaggtttcactatgtgtccagggttaactttgaa 3240
Db 97798 CTAGTGTTTTAAATTTTGTGGAGATGAGGTTTTCATGTATGTTGTCAGGCTAAATCTTGAA 97857
Qy 3241 ctctcggccttaagaacacctcttgctcagcctccacacagtgctaggtattacaagcgtg 3300
Db 97858 CTCCTCGCTTTAAGCAACCCCTCTGCTCAGCCCTCCACAGTGTGATGATTACAAGCGTG 97917

RESULT 3
AC005098 AC005098 230552 bp DNA PRI 07-OCT-2000
LOCUS Homo sapiens BAC clone CTA-350L10 from 7q11.2, complete sequence.
DEFINITION AC005098
ACCESSION AC005098
VERSION AC005098.2 GI:9211526
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 230552)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE 2 (bases 1 to 230552)
AUTHORS Threide,J., Abbott,A., Graves,T., Elliott,G. and Markovic,C.
TITLE The sequence of Homo sapiens BAC clone CTA-350L10
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 230552)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 230552)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 230552)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_RG350L10
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:sgreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelobAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10; actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 222330 to 222775 is a GA rich region. The sequence is not exact but it is believed to be the best representation of this region. The region was sized by PCR from clone DNA at 650 bp. The region corresponds to restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES

source

Location/Qualifiers

1. .230552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-350L10"
/clone_lib="CITB-HS-A"
441. .725
/rpt_family="Alu"
783. .1149
/note="similar to EST BE000775 (NID:g8261008)"
910. .1312
/note="similar to EST R22057 (NID:g776838) yh25f08.sl"
1286. .1716
/note="similar to EST H72453 (NID:g1044269) ys04f07.r1"
1609. .1864
/rpt_family="Alu"
1984. .2271
/note="similar to EST R22113 (NID:g776894) yh25f08.r1"
2337. .2682
/note="similar to EST AW383168 (NID:g6887827)"
2437. .2558
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
2532. .2601
/rpt_family="MIR"
3054. .3367
/rpt_family="Alu"
3240. .3391
/note="similar to EST A1566941 (NID:g452393) tn25b07.x1"
3388. .3845
/note="similar to EST H70120 (NID:g1040326) ys01a09.r1"
3599. .4017
/note="similar to EST H89808 (NID:g1080238) yu82g03.r1"
3741. .3928
/note="similar to EST N56845 (NID:g1200735) yw83e04.r1"
3816. .3932
/note="similar to EST AW523936 (NID:g7166321)"
3819. .3932
/note="similar to EST AW503416 (NID:g7139965)"
3820. .3932
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
3820. .3932
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"

misc_feature 3871. .3932
/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature 3876. .3932
/note="similar to EST T31784 (NID:g613882)"
misc_feature 3889. .4387
/note="similar to EST T79157 (NID:g697666) yd70b07.sl"
misc_feature 4109. .4387
/note="similar to EST A1696334 (NID:g4984234) tw59e05.x1"
repeat_region 4502. .4812
/rpt_family="Alu"
5041. .5087
/rpt_family="Mariner"
5254. .5330
/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature 5265. .5316
/note="similar to EST AW503416 (NID:g7139965)"
misc_feature 5265. .5330
/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature 5285. .5330
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature 5265. .5330
/note="similar to EST AW523936 (NID:g7166321)"
misc_feature 5265. .5330
/note="similar to EST T31784 (NID:g613882)"
misc_feature 5285. .5348
/note="similar to EST AW163398 (NID:g6302431) au94g04.y1"
misc_feature 5284. .5330
/note="similar to EST T30644 (NID:g612742)"
misc_feature 5305. .5330
/note="similar to EST AW993840 (NID:g8254025)"
repeat_region 5750. .6041
/rpt_family="Alu"
5964. .6127
/note="similar to EST AA078486 (NID:g1837960)"
misc_feature 6032. .6370
/note="similar to EST A1566941 (NID:g4525393) tn25b07.x1"
repeat_region 6572. .6878
/rpt_family="Alu"
6637. .6883
/note="similar to EST AW748900 (NID:g7663832)"
misc_feature 6884. .7119
/note="similar to EST AW895654 (NID:g8059859)"
misc_feature 6906. .7117
/note="similar to EST AA601344 (NID:g2434969) no16a03.sl"
repeat_region 6939. .6985
/rpt_family="MIR"
7127. .7218
/rpt_family="L1"
7219. .7425
/rpt_family="Alu"
7426. .7476
/rpt_family="L1"
7510. .7697
/note="similar to EST AA985239 (NID:g3163764) am80d09.sl"
misc_feature 7514. .7697
/note="similar to EST T31784 (NID:g613882)"
misc_feature 7514. .7697
/note="similar to EST AW523936 (NID:g7166321)"
misc_feature 7514. .7689
/note="similar to EST A1700706 (NID:g4988606) we39e12.x1"
misc_feature 7514. .7695
/note="similar to EST T30644 (NID:g612742)"
misc_feature 7514. .7695
/note="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature 7514. .7697
/note="similar to EST AW993840 (NID:g8254025)"
misc_feature 7514. .7697
/note="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature 7565. .7697
/note="similar to EST Z43001 (NID:g570579)"

Query Match

99.0%; Score 3267.6; DB 9; Length 230552;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 3293; Conservative 0; Mismatches 4; Indels 3; Gaps 2;

```
QY 1 tactaaaaatacaaaattagccaggcgtgtgtgycgcacactgtaatacccagctacttgg 60
Db 47036 TACTTAAAAATACAAAATTAGCCAGCGCTGGTGGCGCACACTGTGTAATCCCGAGCTACTTGG 47095
QY 61 gaagctgagcagagagaatcgcttgaactggaagcgagaggttgcagtgcagccgagattc 120
Db 47096 GAAGCTGAGGCGAGGAGAAATGCGTTGAACCTTGAAGCGAGAGGTTGCGAGTGAGCGGAGATT 47155
QY 121 gtgcactgcactcagcctggggcaacaagcgaaacttcgccttcaacaacaataaatta 180
Db 47156 GTGCCACTGCACCTCCACGCTTGGGCAACAAGCGGAACATTCGCTTCAACAATAAATTA 47215
QY 181 agccccagcatgtttggtttcaatctgcagaaactcaacccctcaacccccagagagatcag 240
Db 47216 ACGCCCAAGCATGCTTGGCTTTATCTGCGAGAGCTCAAGCCCTCACCCCGAGGAGATCAG 47275
QY 241 gtccggaccatgagctgaccctggaactcagcaaggatgagttggtgcagccctggcctg 300
Db 47276 GTCCGGACCATGAGCTGACCCCTGAGCTCAGGCAAGGCTGAGTTGGTGCAGCCCTGGGCGTG 47335
QY 301 ctgggagcagagctgcagcagctgcctggggctgagggcgccgcactcaatgaactcat 360
Db 47336 CTGGGAGGCACAGGCTGCACAGGCTGCCCTGGGGCTCAGGCCCCGCCACTCATGAACATCAT 47395
QY 361 gaccttgaatgagctccaaagctctgggctccaggtctccaggtctgaggggagtgggagagag 420
Db 47396 GACCTTGAATGAGCTCCAAAAGCTCTGGGGCTCCCGAGGCTCTAGGGGGAGTGGGAGAGAG 47455
QY 421 aggcctcagcctgtccctgggcatgctgccccctccctcactctttgtcccaaatccct 480
Db 47456 AGGCCTCAGCCTGTCCCTGGGCATGCTGCCCTCCCTCCACACTCTTTGTCCTCCAAATCCCTCT 47515
QY 481 tcctggcaagctgacagtcttaataatcactctgggaaaaactgagtcagccctaaagaa 540
Db 47516 TCCTGGCAAAAGCTGACAGTCTTAATATCACTCTGGAGAAAACCTGAGTCAGCCCTTAAGGAA 47575
QY 541 caattcaatgaaccatttgttacttgatgagatggaaactcaagtctcaactcaaatgtgt 600
Db 47576 CAATTCAAATGAACCATTTGTCTTCTGAGGATTTGAACTCAAGTCTCCTCACTCAAAAGTCTGT 47635
QY 601 gcaatttctgccagctgctcactggccctcaaccacacaccccaagatgagcatcta 660
Db 47636 GCCATTTTCGCCAGCTGTCACTGGCCCTCATCCACACACACCCAGGATGAGCATCTA 47695
QY 661 acgcttgatgcacactcccatgcccggttcattcaactcaattcaattcaattcaactc 720
Db 47696 ACGCTTGATGCACACTCCCATGCCATGCCCGGTTCATTCACTCATTTCAATTCATTCACTC 47755
QY 721 attcattgactcaattcaattcaattcaattcaattcaattcaattcaattcaattcaattc 780
Db 47756 ATTCATTGACTCATTCATTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 47815
QY 781 acgatccaaaattatttgacctgtgtgcccaggaactagatgaggggctggggtctaga 840
Db 47816 ACGATCCAAATATTATGGCTCTGTGTGCCAGGCACATAGTGCAGGGGCTGGGCTAGA 47875
QY 841 gcccctgataacccggttcatagcctagcttctctgggacacacattgtgtaaggggaga 900
Db 47876 GCCCTGATAAACCGGTTCATGCCCTAGCTTTCTCTGGGACACACATTGTGGTAGGGGAGA 47935
QY 901 ctaaaaaataagtcaggccagggcaagggtggtcctcatgctgaatcccgcaactttggga 960
Db 47936 CTAAAAAAATTAAGTCAAGCCAGCGACGGTGGCTCATGTCCTGAATCCCGACACTTTTGGGA 47995
QY 961 ggcgcagggcagtgaaattacactgaggtccagagttcaagaccagcctggcccaactagag 1020
Db 47996 GGCCGAGCGCAGTGAATTAATCTGAGGTCAGAGGTTCAAGACCAGCCTGGCCACATGGAG 48055
QY 1021 aaacccagctcttaattaaaaaataaaaaaataatagccaggtgtgtgtgcacatgcctg 1080
Db 48056 AAACCCAGTCTCTAATT - - AAAAAAAAAAAAAAAAAAATAACCCAGGTGTGTGTGCACATGCTGT 48113
```

```
QY 1081 taatcccagctactcaggagactaacgcaagagaattgcttgaaccacgagagggcaggggt 1140
Db 48114 TAATCCAGAGCTACTCAGAGAGACTAACGCAAGAGAAATTCCTTGAACCCAGGAGGAGAGGT 48173
QY 1141 tgcggtgagccgagatcgccattgcactccagcctgggaaacaagcgagagctccat 1200
Db 48174 TCGGCTGAGCGGAGATCGCGCAATTGCATCCAGCCTTGGGAAACAAGCGAGACTCCAT 48233
QY 1201 ctcaaaaaaataaagattgggagcgagcgagagatcactagagccagtgatttga 1260
Db 48234 CTC - AAAAAAAAAAATGGGAGCGACGAGCGAGGATCACTAGAGCCAGTAGTTTGA 48292
QY 1261 gaccatcctggggaacatagcaggaccctgtctgtaaaaaataataaaaaatttaac 1320
Db 48293 GACCATCTGGGCAACATAGCAGGACCTGTCTGTACAAAAAATTAATAAAAAATTTAAC 48352
QY 1321 cgggcatgttgccacacacccgtagtcctcagctactcagaggtcagggcgagggatcg 1380
Db 48353 CGGGCATGTGTGGCACACACCCGTAGTCCAGCTACTCCAGAGGCTGAGGACGAGGATCG 48412
QY 1381 ctggagccccaggagttggaggtgcagtgaaactgtgtcccccactgcacttaagcctg 1440
Db 48413 CTGGAGCCCGAGGAGTTGGAGGCTGCAGTGAACGTGTGATCCACCACCTGCACCTTAAGCCTG 48472
QY 1441 gataaaaaaagcaagaccctgtctcaataaacaatagcaataataaaaaaataataaa 1500
Db 48473 GATAACAAAAGCAAGACCCCTGTCTCAATAACAATAGCAATAAATAAAGAAAAATTTAAA 48532
QY 1501 tgcaatttgcatgcatcagtgataagctctgcagaaaaagagcgaggaagagctg 1560
Db 48533 TGCATTTTGGCATGCATCAGTGAATAGTGTCTGCAGAAAAGGAGGACGAGGAGCTG 48592
QY 1561 agaagdtatgaggtttgctatgcaatgtgaagttaaaaggaaggcttctcggaagagg 1620
Db 48593 AGAAAGGTATGAGGTTTGCTATGCAATGTGAAGTTATCAAGGAAAGGCTTCTCGGAAGAGG 48652
QY 1621 tgacatttgacagcaagaatggagagagattatgagggagaatggtgaatgggggaac 1680
Db 48653 TGACATTTGACAGAGAAATGGAGGAGATTATGGAGGAAAGATGGTGAATGGGGGGAAC 48712
QY 1681 atggtcaagaccaggaatatggtcaaggggggaagatggtcaaggggagcgagcaaatg 1740
Db 48713 ATGTCGAAGACCCAGGAATATGGTCAAGGGGGGAAGATGGTCAAGGGGACGACGCAATG 48772
QY 1741 caaaggccctgaggcagagcagcttgattcaaccccccaaacccgtggggccctgcagg 1800
Db 48773 CAAAGGCCCTGAGCGAGGAGCAGCTTGATTCAACCCCAAAACCCGCGGGCCCTGCAGG 48832
QY 1801 cgagggaaagacaagtgtaaacccctttcctgtccctgcaggtgtgtgtgaacatgag 1860
Db 48833 CGACGGGAAGGACAAGTGTAAACCCCTTTCTGTCTCCCTGCAGGCTGTGTGAACATGAG 48892
QY 1861 tctgcccatgtttacacccctgcaagcctgaagagtcctcccgaaactgaaagaagaacaa 1920
Db 48893 TCTGCCCATGTTTACACCTTGCAAGCTGAAGAGTCCCCAGAAACTGAAAGAAGAACAA 48952
QY 1921 agcccttctgtacccctccctgcctgcctgtcccgaccgcgacaaaaagcacttcccttt 1980
Db 48953 AGCCCTTTCTGTAGCCTCCCTGCCTGCCTTCCCGACGCGCAAAAAGCGACTTCTCTCTT 49012
QY 1981 ccagtgcatttaagcgccagcctggaagtccagggagcactggagccaccagtcag 2040
Db 49013 CCAGTGATTTAAGGGCGAGCCTGGAAAGTGCAGGGAGCAGCTGGAGGCCACCCAGTATG 49072
QY 2041 ggggacaccttcatccgtcaatcgccctgtcggtcttgagaagcgtctcgtaaccagc 2100
Db 49073 GGGGACACCTTCATCCGTCACATCGCCCTGCTGGGCTTTGAGNAAGCGCTTGTATCCAGC 49132
QY 2101 cagcactatgtgtagtgcgtggtgaggggcatcccggtgggggatacgggagggagacgca 2160
Db 49133 CAGCATTATGTGAGTAGTGTGTGGGGGCATCCCGTGGGGGAATACGGGAGGACACGCA 49192
```


Db 93758 TTCCCTTTGGACCCGCTGGGCCCCAGCCCTTACTGTCCCCGCTGTGTCCCCAGCCAG 93817
QY 2341 gcctcagccttagcaggagtcctcttttctgctccctgccaatgcccagcagccagc 2400
Db 93818 GCCCTAGCCTTAGCCAGGAGTCTCTTCTGTCTCCCTGCCATGGCCAGGCCAGCCAGC 93877
QY 2401 gctctctcaggtccgagggcccatctctccaggaagccttccctgaactagccccagctatca 2460
Db 93878 GCTCTCTCAGTCCGAGGCCCATCTCTCCAGGAAGCCTCCCTGACTAGCCAGCTATCA 93937
QY 2461 gagagtgcctcccaagaaggaggcctgaaactaaagtctctctctcccagctgcc 2520
Db 93938 GAGAGTGGCCCTCCCAAGAGGGAGGCGTGGAATAAGGTCTCTCTCTCCAGCTGCC 93997
QY 2521 tctagtcagtttagctcttctcctcagtagggtgacacatgacaggggccaata 2580
Db 93998 TGTAGTGTCAAGTTAGAGTCTTATCTCTCAGTAGGGTGACACATGACAGGGGCCAATA 94057
QY 2581 gagtctcccatctgttcccaaggaggtggacaaaagcctgctcagacacacaaagtcca 2640
Db 94058 GAGTCTCCCATCTGTCCCCAGGAGGCTGACAAATGCCCTGCTCAGACACACAAGTCCA 94117
QY 2641 ctgggtcccttaatacccataggaagccagggaggaactacattaggaatgaagctt 2700
Db 94118 CTGGGTCCCTTAATCCCATAGGAAGGCCAGGAGAACTACATTTAGGAATTTGAAGCTT 94177
QY 2701 gtatggacatttagtctctatgtgcaagaccttctcttttttttttttttttttttt 2760
Db 94178 GTATGGAAACATTTAGTCTATGTATGCCAAGACCTTCTCTTTTGTATTTTGTGTT 94237
QY 2761 ttgagacagagtccttgatctgttgccagagccagagtgccagtgccagatccagctcac 2820
Db 94238 TTGAGACAGAGCTGTGATCTGTTGCCAGGCCAGAGTGCGAGTGCGCAGATCTCAGCTCAC 94297
QY 2821 tgaacctccgcttccaggttcaactggttctcctgcctcagcctccagagtagttggg 2880
Db 94298 TGAACCTCGGCTTCCAGGTTCAAGTGTCACTGCTCTCCTGCTCCAGCTCCAGAGTAGTTGGG 94357
QY 2881 attacaggtgccaccaccacgcctgctgctaattttttttttttttttttttttttttt 2940
Db 94358 ATTACAGGTGCCACCACACAGCCTGGCTTAATTTTGTATTTTGTATTTTGTAGAGACAGGTTT 94417
QY 2941 caccatgttgccagagctgtctcaactcctgacctcaagtgtacccaccacctgggcc 3000
Db 94418 CACCATGTTGGCCAGACTGTGTCAAACCTCTGACCTCAAGTGATGATCACCCACCTGGGCC 94477
QY 3001 tcccaagagctcgggattacaggcagtagaccacgctgctgctgctgttttttttttttttt 3060
Db 94478 TCCCAAGTGTCTGGGATTACAGGCATGAGCCACCGTGCCTGGCTGTGTTTTTTTGAATGA 94537
QY 3061 ggtctgagtgagtggtgcgatactagttcactgcagcctcaacctcccccagcccaagt 3120
Db 94538 GGTCTGAGGTGCAGTGTGGGATCATAGTTTCATGTCAGACCTCGACCTCCAGGCCCAAGT 94597
QY 3121 gatcctcctgctcagcccttagtagctggtgggtacagcgccacacccaccatgctgg 3180
Db 94598 GATCCTCCTGCCTCAGCCCTTGAGTAGCTGGGCTACAGCGCCACACACCACCATGCTCG 94657
QY 3181 ctggttttttaaaattttttgtggagatgaggtttcaactatattgtccaaggctaatttgaa 3240
Db 94658 CTAGTTTTTAAAAATTTTGTGGAGATGAGGTTTCACTATGTTGTCCAGGCTAATCTTGAA 94717
QY 3241 ctctcggttaagcaacctctggtctcagcctccacagctcaggtattacaagcgtg 3300
Db 94718 CTCTCGGCTTAAGCAACCTCTGGTCTCAGCCTCCACAGTCTCAGGATTAACAAGCTG 94777

RESULT 5
AC004166 AC004166 275197 bp DNA PRI 20-JUL-2000
LOCUS Homo sapiens Chromosome 7 BAC clone 239c10, complete sequence.
DEFINITION AC004166
ACCESSION AC004166
VERSION AC004166.12 GI:8887011

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 275197)
Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
Homo sapiens Chromosome 7 BAC Clone 239c10
Unpublished
2 (bases 1 to 275197)
Burian,D.M. and Roe,B.A.
Direct Submission
Submitted (21-FEB-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 275197)
Burian,D.M., Ren,Q., Meadows,S., Huang,E., Korenberg,J. and
Roe,B.A.
Direct Submission
Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 275197)
Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
Direct Submission
Submitted (20-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jul 1, 2000 this sequence version replaced gi:8779478.
FEATURES
Location/Qualifiers
source
1..275197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="239c10"
BASE COUNT 70874 a 66648 c 66737 g 70938 t
ORIGIN
Query Match 99.0%; Score 3265.6; DB 9; Length 275197;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3292; Conservative 0; Mismatches 4; Indels 4; Gaps 2;
QY 1 tactaaaaatcacaaaattagccagggctggtgcygcacacctgtaatcccagctacttgg 60
Db 33415 TACTAAAATTACAAAATTAGCCAGGCGTGTGGCGCACACCTGTAATCCCAGCTACTTGG 33474
QY 61 gaagctgaggcaggaggaatcgcgttgaacctgggaaggcagaggttgagtgagccgagatt 120
Db 33475 GAAGCTGAGCGACGAGAAATCGCTTGAACCTGGGAAGGCAGAGGTTCAGTCGAGCGAGATT 33534
QY 121 gtccactgcactccagcctgggcaacaagaggaacttcgttcaacaataaatta 180
Db 33535 GTGCCATGTCACATCCAGCTTGGGCAACAAGAGCAAACTTCGCTTCAACAATAAATA 33594
QY 181 acgcccagcgtgtcttggtcttctatctgcagacactcaacctcagccagagatcag 240
Db 33595 ACGCCAGCATGCTTGGCTTTCATCTGCCAGACCTCAACCTCACCCTCCAGGAGATCAG 33654
QY 241 gtcggaccatgagctgacctggactcaggaagggtgagttggtgagccctggcctg 300
Db 33655 GTCGGACCATGAGCTGACCTGAGCTCAGGAGGGGTGAGTTGGTGCAGCCCTGGCCTG 33714
QY 301 ctggaggcacaggtctcagcaggtcctgggctgagggccgcccactcataaactcat 360
Db 33715 CTGGAGGCACAGGCTCAGCAGGCTGCCCTGGGGCTGAGGCCGCCACTCATGAACTCAT 33774
QY 361 gacctgaatgagctccaaaaagctctgggctccagagctctaggggagtgaggagagag 420
Db 33775 GACCTTGAATGAGCTCCAAAAGCTCTGGGCTCCAGGCTCTAGGGGAGTGGGAGAGAG 33834
QY 421 aggcctcagcctgtccctgggcatgctgccccctctcaactcttttccccaaatccct 480
|||||

Query Match				
Best Local Similarity 63.8%; Score 2106.8; DB 9; Length 8119;				
Matches 2133; Conservative 0; Mismatches 7; Indels 2; Gaps				
Qy	1161	ccattgcactccagctgggaacaagacgagactccatctc-aaaaaaaaaaaagtg	1219	
Db	1		60	
Qy	1220	ggaggcagaggcaggagatcactagagccagtagtlttgagaccatctctgggcaacata	1279	
Db	61	GGAGGCAGAGCAGGAGGATCACTAGAGGCCAGTAGTTGAGACCATCTCTGGGCACATA	120	
Qy	1280	gcaggacctgtctgtacaaaaaattaaaaaatttaaccgggcatctgtgtgcacacac	1339	
Db	121	GCAGGACCTGTCTGTACAAAAAATTAAAAAATTTAAACCGGCATGTTGGCACACAC	180	
Qy	1340	ccgtagtccagctactccagaggtgagcgaggagatcgtggagcccaggagtctgga	1399	
Db	181	CCGTAGTCCCAGCTACTCCAGAGCTTGAGCGAGGAGATCGCTGGAGCCCAAGGAGTTGGA	24	
Qy	1400	ggctgcagtgaactgtatccaccactgcacttaagctctggtataacaagaacaagacct	1459	
Db	241	GGCTGCAGTGAATGTGATCCCAACCACCTGCGCTTAGCGCTGGATPAACAGCAGACCCCT	300	
Qy	1460	gtctcaataacaatgcaataataaagaaaaattaaatgcaatttgcgtatgcata	1519	
Db	301	GTCTCAAAATAACAATAGCAATAATAATAAGAAAAAATTAATGCAATTTGCGATGCATCA	360	
Qy	1520	gtataagtgctctgcagaaaaaggagcgaggaagagcctgagaaaggtatgaagtttgc	1579	
Db	361	GTGATAAGTGTCTCTGCAGAAAAGGAGGCGAGGAAGAGCTGAGAAAGGTATGAGGTTTGC	420	
Qy	1580	tatgcaatgtgaagtattcaagaagaagctcttcgggaagaggtgcacatttgcagcagagaaa	1639	
Db	421	TATGCAATGTGAAGTTATCAAGGAAGGCTTCTCGAGAGAGGTGCATTTGAGCAGAGAAA	480	
Qy	1640	tggaggaggtttatggagggaagatgtgtaattgvgggggaacatggtcgaagccaggaaata	1699	
Db	481	TGGAGGAGAGTTATGGAGGGGAAGATGGTGAATGGGGGGAACATGCTCAAGACCAGGAATA	540	
Qy	1700	tgttcaagggggaagaatggttcaaggggacacgacaattgcaagggccttgaggcagga	1759	
Db	541	TGGTCAAGGGGGGAAGATGGTCAAGGGGACGCGCAATGCAAGAGGCCCTTGAGGCGAGGA	600	
Qy	1760	gcagcttgattcaccccccacaaaccgtggggccgtgcaggcgagggaggaacaagtgt	1819	
Db	601	GCAGCTTGATTACCCCCCAAAACCCGTGGGGCCCGTCGAGCGGAGGGGAGACAGTGT	660	
Qy	1820	aaacctttctctgttccttcgcaggtgtgtgtgaacatgagctgtgccatgtttacacc	1879	
Db	661	AAACCCCTTTTCCCTTGTCCCTTGCAGGTGTGTGTGAACATGAGTCTGCCCATGCTTTACACC	720	
Qy	1880	tgaacctgaaagttccccagaactgaaaagaagaagaccccttctgtacctctcc	1939	
Db	721	TGCAAGCCTGAAGAGTCTCCAGAACTGAAAAGAAAGAACGCCCTTTCTGTACCCCTCC	780	
Qy	1940	ctgccctctgccgcaccgcgcacaaagcgacttctctcttccagtgcatttaaggcgca	1999	
Db	781	CTGCCCTCTGTCCCGACCGGACAAAGGCACTTCTCTTCCAGTGCATTTAAGGCCECA	840	
Qy	2000	gcctggaagtgcaggaggacatgtaggccaaccagtcgatggggacaccttcatccgtc	2059	
Db	841	GCCTGGAAGTGCACGAGGACACTGGAGGCCACCCCACTCATGGGGACACCTTCATCCGTC	900	
Qy	2060	acatcgccctgtcgtgttttgagaagcgtcttgttaaccagccagcactatgtgagtacct	2119	
Db	901	ACATCGCCCTTGTCTGGGCTTTTGAGAAGCGCTTCGTACCCAGCCAGCACTATCTGAGTAGCT	960	
Qy	2120	ggttgaggggcat-cccggtggggggaatccggagggacagcacgcccacccttgcagtcc	2178	
Db	961	GGTGAGGGGCACTCCCGTGGGGGGAATACCGGAGGGGACAGACNAGGCCACACCTTCGAGTCC	1020	
Qy	2179	caggggcaaacagctccagtgaggactaacggggcgaggggtcttgtggcaaccttggtccctgg	2238	

[illegible]

Db 2101 CCTGTGCTCAGCCTCCACAGTGTAGGATACAGCGTG 2142

RESULT 9
HS47PHOX01
LOCUS
DEFINITION Homo sapiens p47-phox (NCF1) gene, exons 1-5.
ACCESSION U57833
VERSION U57833.1 GI:2754709
KEYWORDS
SEGMENT
SOURCE 1 of 3
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8131)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
MEDLINE 97474758
REFERENCE 2 (bases 1 to 8131)
AUTHORS Chanock,S.J., Roesler,J., Hopkins,J., Lee,P., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Genomic Structure and the Identification of Multiple Polymorphisms
in the p47-phox Gene
Unpublished
REFERENCE 3 (bases 1 to 8131)
AUTHORS Gorlach,A., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1996) P. Lee, Molecular and Experimental
Medicine, Scripps Research Institute, 10550 North Torrey Pines Rd,
La Jolla, CA 92037, USA
FEATURES
source
1. .8131
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="p1 clone 42"
/map="7q11.23"
877. .948
/gene="NCF1"
/number=1
4143. .4223
/gene="NCF1"
/notes="this is defined as the wildtype gene since it
contains GTGT at the beginning of exon 2"
/number=2
5957. .6032
/gene="NCF1"
/number=3
6132. .6297
/gene="NCF1"
/number=4
7657. .7712
/gene="NCF1"
/number=5
BASE COUNT 1835 a 2214 c 2088 g 1994 t
ORIGIN

Query Match 63.8%; Score 2105.8; DB 9; Length 8131;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2132; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy 1161 ccattgcactccagcctgggaacagagcagagcactccatctcaaaaaaaagtgg 1220
|||||
Db 1 CCATTGACCTCCAGCCTGGGAACACAGAGGAGACTCCATCTC-AAAAAATAAGTGG 59
|||||
Qy 1221 gaggcagagcagagagatcactagagccagtagtttgaccatcctgggaacatag 1280
|||||

QY	2360	agtcctctttctgtctcccctgccatgcccaggccaggccagcgtctctcaggccagggc	2419
DB	1200	AGTCCCTTTTCTGCTCCCTGCGCATGGCGAGCGACGCCAGCGCTCTCTCAGGTCCGAGGC	1259
QY	2420	ccactctccaggaaagccttccctgactagcccagctatcacagagatggccctcccaaga	2479
DB	1260	CCACTCTCCAGGAAGCCTTCCTGACTAGCCAGCTATCAGAGATGGCCCTCCCAAGA	1319
QY	2480	gggagccctggaaaactaaagctctctctctcccagctgctgtagtgctcagttagagtc	2539
DB	1320	GGGAGCCCTGGAAACTAAAGCTCTCTCTCTCCCCAGCTGCCTGTAGTGTACAGTTACAGTC	1379
QY	2540	ttatcctctcagtagggtgacacacatgaacaggggccaaatagagtcctcccactctcccc	2599
DB	1380	TTATCTCTCTCCAGTAGGCTGACACCATGACAGGGCGCAATAGAGTCCTCCCATCTCTCC	1439
QY	2600	caaggaggctggacaaaatgcctgctcagacacacaagtccactgggtcccttaatacccat	2659
DB	1440	CANAGGAGCTGGACAAATGCCCTGCTCAGACACACAGTCCACTGGGTGCCCTTAATCCCAT	1499
QY	2660	aggaagccaggaggaaactacatttagaaattgaagcttgtatggaacatttagtctct	2719
DB	1500	AGGAAGCCCAGGAGGAACATACATTTAGGAAATTTGAAGCTGTGATGGAACATTTACTCCT	1559
QY	2720	atgfgcaagaaccttctcttttttggttattttttgtgttttggtttttgagacagagctctgac	2779
DB	1560	ATCTGCCAAGACCTTTCTCTTTTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1619
QY	2780	tgctgccaggccagatgcagtgagcagatctcagctcactcaactcagcctccgcttccag	2839
DB	1620	TGTTGCCCAGGCCAGATGTCAGTGGCAGATCTCAGCTCACTGAAACCTCCGGCTTCCAG	1679
QY	2840	gttcaactggttctctgcctcagcctccagtagtgattgggattacaggtgcccacacc	2899
DB	1680	GTTCAACTGGTTCTCCTGCCTCAGCCTCCAGAGTAGTTGGGATTACAGGTGCCACCACC	1739
QY	2900	acgcctggcctaatttttgbatttttagtagacagaggtttcaacctgttggccagactg	2959
DB	1740	ACGCCTGGCTAAATTTTGTATTTTGTAGTAGAGACAGGGTTTCAACATGTTGGCCAGACTG	1799
QY	2960	gtctcaactcctgacctcaagtgatccaccacctgggacctcccgaagtgtcgggattta	3019
DB	1800	GTCTCAAACTCTGTACCTCAAGTGATCCACCACCTGGGCTCCCAAGTGTGGGATTTA	1859
QY	3020	caggcatgagccaccgtgcctggcctgttttttgaattgaggtctggagtcagtggtg	3079
DB	1860	CAGCATGAGCCACCGTGCCTGGCCTGTTTTTTTGAATGAGGTCTGGAGTGCAGTGTGTG	1919
QY	3080	cgatcatagttcactgcagcctcaacctcccaggcccaagtgtatctctgcctcagccc	3139
DB	1920	CGATCATAGTTTCACTGCAGCCTCAAGCTCCAGGCCCAAGTGATCGCGCTGCCTCAACCC	1979
QY	3140	cttgagtagctgggctcacaggccacacacacatgctggctagtttttaaaatttttg	3199
DB	1980	CTTGAGTAGCTGGGCTACAGGCGCACACCACCATGCTGGCTAGTTTTTAAATTTTTTG	2039
QY	3200	tggagtagggtttcacatagtgtgcaggctaatcttgaaactcctcggtcttaagcaacc	3259
DB	2040	TGGAGATGAGTTTCACTATGTTGTCCAGGCTAATCTTGAACCTCTCGGCTTTAAGCAACC	2099
QY	3260	ctctgctcagcctcccacagtgctagattacaacgctg	3300
DB	2100	CTCTGTCTCAGCCCTCCACAGCTGCTAGGATTACAGCGTG	2140
RESULT 10			
AF003533			
LOCUS			
DEFINITION			
Homo sapiens cytosolic phagocyte oxidase protein (p47phox) gene,			
promoter region and partial cds.			
ACCESSION			
AF003533.1			
VERSION			
AF003533.1 GI:2197080			
PRI			
29-JUL-1997			

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 3074)
TITLE Li, S.-L., Valente, A.J., Zhao, S.J. and Clark, R.A.
JOURNAL PU.1 is essential for p47(phox) promoter activity in myeloid cells
MEDLINE J. Biol. Chem. 272 (28), 17802-17809 (1997)
REFERENCE 9736227
AUTHORS 2 (bases 1 to 3074)
TITLE Li, S.-L., Valente, A.J., Zhao, S.J. and Clark, R.A.
JOURNAL Direct Submission
SUBMITTED (09-MAY-1997) Medicine, U. of Texas Hlth Sci Ctr, 7703
FLOYD CURTIS Drive, San Antonio, TX 78284, USA
FEATURES
source
1..3074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23"
/note="Cloned in PromoterFinder Kit from Clontech (Palo Alto)"
promoter
1..3050
/gene="p47phox"
3072..>3074
/gene="p47phox"
3051..>3074
/gene="p47phox"
3051..3071
/gene="p47phox"
3072..>3074
/gene="p47phox"
/note="cytosolic phagocyte oxidase protein"
/codon_start=1
/product="p47phox"
/protein_id="AA864186.1"
/db_xref="GI:2197081"
/translation="M"
BASE COUNT 849 a 763 c 842 g 620 t
ORIGIN
Query Match 59.9%; Score 1978; DB 9; Length 3074;
Best Local Similarity 99.1%; Pred.No. 0;
Matches 202; Conservative 0; Mismatches 10; Indels 8; Gaps 3;
QY 1 tactaaatacaaaattagccaggcgtggtgcgacacctgtaatccacgactctgg 60
Db 1043 TACTAAAATACAAATTAGCAGCGCTGTGGCGCACACTATAATCCAGCTACTTGG 1102
QY 61 gaagctgagcagagagaatcgcttgaaacctggaagcgagaggttcagtagccagatt 120
Db 1103 GAAGCTGAGCAGAGAGATCGCTTGAACCTGGAAGGAGAGGTTGCAGTGAGCGAGATT 1162
QY 121 gtgcactgcactccagcctgggcaacaagagcgaaacttcgcttcaacaaataaatta 180
Db 1163 GTGCCACTGCACCTCCAGCTGGGCAACAAGAGCGAACTTCGCTTCAACAATAAATTA 1222
QY 181 acgcccagcatgtcttggctttcatctgccagacctcaacccctcaacccccaggagatcag 240
Db 1223 ACGCCAGCATGTCTTGGCTTTCATCTGCCAGAGCTCAACCCCTCACCCCCAGGAGATCAG 1282
QY 241 gtccgagcatgagctgacctggactcaagcgaagggtgagttggtgcagccctggcctg 300
Db 1283 GTCCGAGCACAGAGCTGACCTGGACTCAGGCAAGGCTGAGTTGGTGCAGCCCTGGCGCTG 1342
QY 301 ctgggagcagagctgcagcagctgccttgggctgagggcgcgcactcatgaactcat 360
Db 1343 CTGGGAGCACAGGCTGCAGAGGCTGCCCTGGGGCTGAGGCCGCCCGCCACTCATGACTCAT 1402
QY 361 gacctgaatgagctccaaagctctggcctccagagctctagggggagtgaggagagag 420


```

exon
1. >1965
/gene="NCF1"
/pseudo
/note="contains 5' UTR"
/number=1
BASE COUNT      551 a   503 c   517 g   394 t
ORIGIN
Query Match      58.9%;   Score 1943.4;   DB 9;   Length 1965;
Best Local Similarity 99.6%;   Pred. No. 0;
Matches 1958;   Conservative 0;   Mismatches 6;   Indels 1;   Gaps 1;

Qy  74  gagaatcgctgaacctggaaaggcagaggtgtcagtgagccgagattgtgccactgcaact 133
      |||||||
Db   1  GAGAAATCGCTTGAACCTCGGAAGGAGAGGTTTGCAGTGAAGCCGAGATTGTGCCACTGCACT 60

Qy  134  ccagctggcgcaacaagcgaaacttcgcttcaaacaaataaatttaacgccagcatgt 193
      |||||||
Db   61  CCAGCTGGGCAACAAGACGCAAACTTCGCTTCAACAATAATAATTAACGCCACAGCTGT 120

Qy  194  cttggctttcatctgcagacactcaacccctcacccccaggagatcaggtccggaccatga 253
      |||||||
Db  121  CTTGGCTTTTCATCTGCACAGACTCAACCCCTCACCCCCAGGAGATCAGTCCGGACCACGA 180

Qy  254  gctgacctgactcaaggcaagggtgaattgtgtcagccctggcctgctggaggacag 313
      |||||||
Db  181  GCTGACCCCTGGACTCAGGCAAGGGTGAATTGGTGCAGCCCTGGCCTGCTGGAGGCGACAG 240

Qy  314  gctgcagcagctgcttgggctgagggcccgccactcatgaactcatgaccttgaatgag 373
      |||||||
Db  241  GCTGCAGCAGGCTGCCGTGGGGCTCAGGCCACACACTCATGAACATCATGACCTTTGAATGAG 300

Qy  374  ctccaaaagctctgggctcccaggtcctagggggagtgggagagagagggcctcagcctg 433
      |||||||
Db  301  CTCCAAAAAGCTCTGGGGCTCCACAGCTCTTAGGGGGAGTGGGAGAGAGAGGCTTCAGCCTG 360

Qy  434  tccctgggcatgctgccccctcctcaacctttgttcccaaatcccccttcctggcaagct 493
      |||||||
Db  361  TCCCTGGGCATGCTGCCCCCTCCACACTCTTGTGCCAAATCCCTTCCTGGCAAAAGCT 420

Qy  494  gacagctcttaatactactctggagaaactgagtcagcctcaaggaaacaaatcgaatgac 553
      |||||||
Db  421  GACAGCTTAAATATCACTCTGCGAGAAACTCAGTCAGGCCCTTAAGGAACAAATTCAATGNAC 480

Qy  554  cattgttacttgaggatbgaactcaagtcctcaactcaaaagtctgtgccatttctgcc 613
      |||||||
Db  481  CATTGTGCTTACTTGAGGATTGAACTCAAGTCTCACCTCAAAAGTCTGTGCCATTTTCGTGCC 540

Qy  614  cagctgcactggccctcatccagacacacccaaggatgagcatctaaagcgttcgatgca 673
      |||||||
Db  541  CAGCTGTGCATGGGCCCTCATCCACACACACCAAGGATGAGCATCTAACGGTTGCATGCA 600

Qy  674  cactcccatgcgcggttcattcactcattcatcattcattcactcattcattcactgactca 733
      |||||||
Db  601  CACTCCCATGCCCCGGTTCAATTCATCTCATTCATTTCATTCATTCATTCATTCATTCATCA 660

Qy  734  ttcatcattcaactcaactcaattcatcactcaagtgaatgtgtcagtcacgcatcccaatat 793
      |||||||
Db  661  TTCAATTCATTCACATCACTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 720

Qy  794  ttatggcctctgtgccagcactagatgaggggctgggctagagccctgataacc 853
      |||||||
Db  721  TTATGGCCCTCTGTGTGCCAGGCATAGATGAGGGGCTGGGGCTTAGAGCCCTCATTAACC 780

Qy  854  cggtcatgccctagctttcctgggacacacattgtgttaaggggagacataaaaaataa 913
      |||||||
Db  781  CGGTCAATGCCCTAGCTTTCCTGGGACACACATTGTGTTAAGGGGAGACTAAAAAATAA 840

Qy  914  gtcaggccaggacagggtggctcatctgcctgaaatccacgacctttgggaggccgagtgag 973
      |||||||
Db  841  GTCAGGCCAGCAGGTTGGTTCATTCGCTTGAATCCAGACATTTGGGAGGCCGAGGCCAGT 900

```

Qy	974	gaattac	gtgagtc	gaggttca	agaccagc	ctggtg	ccaaacat	tggaagaa	accagg	ttctct	103						
Db	901	GAATT	ACCT	GAGTC	AGGAGTT	CAAGAC	CAGCCTG	GCCAA	CATGG	AAACCCAG	TCCTCT	960					
Qy	1034	aattaaaa	aaaaaaaa	aaattag	ccagg	tgtt	gggcaca	tgccct	gtat	atccag	ctac	1093					
Db	961	AA	TTAAAA	AAAA	AAAAA	TTAC	CCAG	GTGTG	TGGC	ACATGC	CTGTAA	TCC	AGGTAC	1020			
Qy	1094	tcagag	actaac	gcgaag	agaatt	ctctga	accagag	gcagag	gttc	ggtg	agccga	1153					
Db	1021	TC	AGGACTAC	CGCAAG	AGAAT	TGCTT	GA	CCAG	CAGAG	CAGAG	GTTCGG	TGAG	CCGA	1080			
Qy	1154	gatc	gcgc	cat	ttgc	actccag	ctggtg	aaacaag	agcag	agctcca	ctctc	-	aaaa	aaaa	1212		
Db	1081	GAT	CGGCCAT	TGC	ACTC	CAGCCT	TGG	AAACA	AGAG	CAGAC	TCCAT	CTCA	AAAA	AAAA	1140		
Qy	1213	aaa	agtgg	aggcag	agggcag	gagga	gtacta	gaag	gccag	tag	ttt	ga	agacc	atc	ctgg	1272	
Db	1141	AA	AGTGG	AGGC	CAGAG	GAGG	AGAT	CAC	TAG	AGCC	AGTAC	GT	TGAG	ACCAT	CTCTGG	1200	
Qy	1273	caac	atag	cagag	ccctg	ctctg	tac	aaaa	aaaa	att	aa	cccc	ggc	atg	gtg	1332	
Db	1201	CA	ACATAG	CAGAC	CCCTG	CTGT	TAC	AAAA	AAAT	TAAAA	AAAT	T	AA	CCGG	CATGTGG	1260	
Qy	1333	ca	ca	ccg	t	ag	t	c	c	a	g	t	c	c	a	1392	
Db	1261	CAC	ACAC	CCG	TAG	TCC	AGC	TACT	CC	AGG	CTG	AGC	AGG	ATC	GTG	AGC	1320
Qy	1393	ag	ttg	ag	g	ctg	cag	tga	act	gtga	tccc	acc	actg	ca	cttg	ga	1452
Db	1321	AG	T	TGG	AGG	CTC	AGT	GA	CTGT	GAT	CC	ACC	ACT	CG	CTTAA	G	1380
Qy	1453	ag	ac	c	c	t	g	t	c	a	a	a	a	a	a	a	1512
Db	1381	AG	AC	CC	TGT	CTCA	AA	TAA	CA	ATAG	CA	AT	A	TAA	G	AAAA	1440
Qy	1513	tg	cat	cag	tga	ta	ag	tgc	tctc	g	a	a	a	a	a	a	1572
Db	1441	TG	CAT	CAG	TGAT	AGT	GCTCT	GC	AG	AAAA	AGG	AGG	CAG	GA	AGG	TATGA	1500
Qy	1573	ggt	t	g	t	c	t	a	a	g	a	g	g	t	c	t	1632
Db	1501	GCT	TTT	GCTAT	GCAAT	TG	AA	GT	TAAT	CA	AG	AG	GC	TTCT	CG	GAAG	1560
Qy	1633	ag	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1692
Db	1561	AG	AG	AAAT	GG	AG	GAG	AGT	TAT	TG	GAG	GGA	ATG	GT	GAAT	TG	1620
Qy	1693	ag	a	a	a	a	a	a	a	a	a	a	a	a	a	a	1752
Db	1621	AG	GAAT	T	GGT	CA	AGG	GGA	AA	GAT	GGT	CA	AGG	GAC	GCG	CAAT	1680
Qy	1753	gg	cag	ag	cag	ctt	gatt	cac	cccc	caaa	cccg	tgg	gccc	gtgc	ag	gc	1812
Db	1681	GG	CAG	GAG	CAG	CTTG	ATT	CAC	CCCC	AAAA	CCCG	TGG	GG	CCCG	TGC	AGG	1740
Qy	1813	ca	ag	t	g	a	a	a	a	a	a	a	a	a	a	a	1872
Db	1741	CA	AGT	GT	AA	AC	CC	TTT	CT	CT	CC	CTG	CAG	GT	CTGT	G	1800
Qy	1873	ta	c	c	c	c	t	g	a	a	a	a	a	a	a	a	1932
Db	1801	TAC	AC	CC	CTG	CAG	CC	TGA	GAG	T	CC	CC	AG	AA	CTG	AA	1860
Qy	1933	a	c	c	c	c	t	g	c	c	c	c	c	c	c	c	1992
Db	1861	AC	CC	T	CC	CTG	CC	C	GAC	CCG	CA	AA	A	AG	C	ACT	1920
Qy	1993	ag	g	c	a	g	c	c	t	g	a	a	a	a	a	a	2037
Db	1921	AG	GC	C	AG	C	T	G	A	G	T	G	C	A	G	G	1965

```
RESULT 12
HS47P41S01
LOCUS HS47P41S01 1569 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox pseudogene, clone p41, exon 1.
ACCESSION U61241
VERSION U61241.1 GI:2754725
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1569)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
REFERENCE 2 (bases 1 to 1569)
AUTHORS Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
Characterization of the genomic structure of the p47-phox gene
Unpublished
REFERENCE 3 (bases 1 to 1569)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
Direct Submission
Submitted (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source
1..1569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P41"
/chromosome="7"
/map="7q11.23"
/note="Human genomic p1 clone p41 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2."
<878..949
exon
/pseudo
/number=1
BASE COUNT 411 a 425 c 435 g 298 t
ORIGIN

Query Match 47.0%; Score 1552.2; DB 9; Length 1569;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1565; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1161 ccattgactccagcctgggaacaaagagcgagactccatctcaaaaaaagaattgg 1220
|||||
Db 1 CCATTGCACTCCAGCCCTGGGAACAAGAGCGAGACTCCATCTCAAAAAAAGATGG 60
|||||

Qy 1221 gagcacagcagagatcactagagccagtagtttgagaccatctctgggcaacatag 1280
|||||
Db 61 GAGCAGAGCGCAGGAGGATCACTAGAGCCAGTAGTTTGAGACCATCTCTGGGCAACATAG 120
|||||

Qy 1281 caggaccctgtctgtacaaaaaattaaaaaaatttaaacccgggcatgtgtggcacacc 1340
|||||
Db 121 CAGGACCTGTCTGTACAAAAAATTTAAAAAATTTTAAACCGGCGCATGTTGGCACACACC 180
|||||

Qy 1341 cgtagtccagctactccagagctgaggcagggagtagctgagccaggaattggag 1400
|||||
Db 181 CGTAGTCCCACTACTCAGAGGCTGAGGAGGAGGATCGCTGGAGCCCAAGAGTTGGAG 240
|||||

Qy 1401 gctcagtgaactgtatccaccactgcacttaagcctgatacaaaagcaagaccctg 1460
|||||
Db 241 GCTCAGTGAACACTGTGATCCACCACCTGCGCTTAAGGCTGGATTAACAGGCAAGACCCCTG 300
|||||

Qy 1461 tctcaataacaaatagcaataataaagaaaaattaaatgcaatttgcgagcatcag 1520
|||||
```

```
Db 301 TCTCAATACATAGCAATAATAATAAGAAAAATTAATGCAATTTGCGATGCATCAG 360
|||||
Qy 1521 tgataagctctctgcagaaaaaagagagcgagagagctgagaaaaagtagagtttgcct 1580
|||||
Db 361 TGATAAGTGTCTCTGCAGAAAAAAGAGGAGCAGAGAGCTGAGAAAGTATCAGGTTTGGCT 420
|||||
Qy 1581 atcaaatgtgaagtattcaagggaaggctctcgaaagagtgacatttgagcagagaat 1640
|||||
Db 421 ATGCAATGCTGAAGTTATCAAGGAAGGCTTCTCGAAGAGGTGACATTTGACAGAGAAAT 480
|||||
Qy 1641 ggagagagattatggaggggaagtgtgtaattgggggaacatggtcaagaccaggaatat 1700
|||||
Db 481 GGAGGAGAGTTATGAGCGGAAGATGTTGAATGGGGGAACATGTTGTAAGACCAGGAATAT 540
|||||
Qy 1701 ggtcaaggggggaaaagtgtctcaaggggagcagcagcaaatcaaaagggccctgagggcaggag 1760
|||||
Db 541 GGTCAAGGGGGGAAAGATGTTCAAGGGGAGCGCAGCAATGCAAAAGGCCCTTGAGCAGGAG 600
|||||
Qy 1761 cagcttgattcaccccccaaaacccgtggtggccgtgcagcgacggggaagacaagtga 1820
|||||
Db 601 CAGTTTGATTCAACCCCAAAACCCGTCGGGGCCGTGCGAGCGAGCGGAAGGACAAGTGTA 660
|||||
Qy 1821 aaccttttctctgtccctgcaggtgtgtgtaacatgagctctgcccattgtttacacct 1880
|||||
Db 661 AACCTTTTCTCTTCTGCTCCTGAGGTGTGTGTAACATGAGTCTGCCCATGTTTACACCT 720
|||||
Qy 1881 gcaagcctgaagagtcctccagaaaactgaagagaagaacaaagcccttctgtacccctcc 1940
|||||
Db 721 GCAAGCCTGAAGAGTCCCCAGAAACTGAAAGAAAGAAAGCAAGCCCTTCTGTACCCCTCC 780
|||||
Qy 1941 tgccctcttcccagccgcagcaaaaagcactctctcttccagtgcatttaagcgacag 2000
|||||
Db 781 TGCCTCTGTCTCCGACCGGACAAAGGACTTCTCTTCCAGTGCATTTAAGGCGCAG 840
|||||
Qy 2001 cctggaagtgccagggagcactggaggccacccactgcatgggggacaccttcatccgtca 2060
|||||
Db 841 CCTGGAAGTGCAGGAGGACACTGAGGAGCCACCCAGTATGGGGGACACCTTTCATCCGTCA 900
|||||
Qy 2061 catgcctctctggtcttgaagagcctctgtaccacagcagcactatgtgagtactg 2120
|||||
Db 901 CATCGCCCTGCTGGGCTTTGAGAAGCGCTTCGTACCCAGCCAGCACATATGTGAGTAGCTG 960
|||||
Qy 2121 gtgaggggcat-cccgtggggggaatcacgggagggagacagcacggccaccttgcagtc 2179
|||||
Db 961 GTGGAGGCGCATCCCGTGGGGGGAATACGGGAGGAGACAGCACGCCACCTTGCAGTCCC 1020
|||||
Qy 2180 agggccaaacagctccagtgaggactaacggggcagggtcttgggacacctggctcctgg 2239
|||||
Db 1021 AGGCCCAACCAAGCTCCAGTGAGGACTTAACGGGGCAGGGTCTTTGGGCACTTGGTCCCTGGT 1080
|||||
Qy 2240 ctttgagcctggatctaccctctgatccctgggaagacagttcccttggaccgcctctg 2299
|||||
Db 1081 CTTTGAGCCTGGATCTACCCCTCTGATCCCTGGGAAGACAGATTCCCTTGGACCCGCCCTG 1140
|||||
Qy 2300 ggcctccagcccttactgtcccgcctgtgtccccagccagggccctcagccttagccagg 2359
|||||
Db 1141 GGCCCCAGCCCTTTACTGTCCCCGCCCTGTGTCCCCAGCCAGCGCCCTCAGCCCTTAGCCAGG 1200
|||||
Qy 2360 agtctcttttctcctcctgccaatggccaggcagcccaagcgctctctcaggtccaggcc 2419
|||||
Db 1201 AGTCCTCTTCTGTCTCCCTGCCATGGCCAGCGACGCCGCTCTCTCAGGTCCGAGGC 1260
|||||
Qy 2420 ccactctccaggaagccttccctgactagccagctatcacagagtgggccctcccaaga 2479
|||||
Db 1261 CCATCTCTCCAGGAAGCCTTCCCTTGACTAGCCCAAGCTATCAGAGAGTGGCCCTCCCCAAGA 1320
|||||
Qy 2480 ggagggcctggaaaaactaaagctctctctctccacagctgcctgtagtgcagttagagtc 2539
|||||
Db 1321 GGGAGGCCGTGGAACCTAAAGCTCTCTCTCCCCAGCTGCGCTGTAGTGTCAAGTAGAGTC 1380
|||||
Qy 2540 ttatcctctccagtagggtgacacatgacaggggccaatagagtcctcccatctgtccc 2599
|||||
```


[illegible][illegible]

```
Db 901 ATTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGACTGGTCTCAAACTCTTGACCTC 960
Qy 2979 aagtgatccaccacccctgggctcccaagtgctgggatcacagcatgagccacccgtgc 3038
Db 961 AAGTGATCCACCCACCTTGGGCTCCCAAGTGTGGGATTCAGGCGATGAGCCACCGTGC 1020
Qy 3039 ctggcctgttttttgaatagagctcgtggagtcagtggtgctgcatcatagttcactgcag 3098
Db 1021 CTGCGCTGTTTTTTGAAATGAGTCTGGAGTGCAGTGGTGGGATCATAGTTCACCTGAG 1080
Qy 3099 cctcaacctccagcccaagtgatctcctcgtcgtcagcccttgtagtgcgtgggctac 3158
Db 1081 CCTCAAGCTCCAGGCCCAAGTGATCGCCTCGCTCAGCCCTTGAGTAGCTGGGGCTAC 1140
Qy 3159 aggcgcacacaccatgcctggctagtttttaaaatttttgtagatgaggtttcacta 3218
Db 1141 AGGTGACACACCACCATGCTGGCTAGTGTAAAAATTTTGTGGAGATGAGGTTTCACTA 1200
Qy 3219 tgtgtccaggctaattttaactcctcgtccttaagcaacctcgtgtcagcctccca 3278
Db 1201 TGTGTCCAGGCTAATCTGAACCTCTCGGCTTAAGCAACCTCTGCTCAGGCTCCCA 1260
Qy 3279 cagtcgtaggattacagcgctg 3300
Db 1261 CAGTGTAGGATTACAAGCGTG 1282

RESULT 14
HS47P40S01 7258 bp DNA PRI 07-JAN-1998
LOCUS Homo sapiens p47-phox (NCF1) pseudogene, clone P40, exons 1-5.
DEFINITION U61238
ACCESSION U61238.1 GI:2754721
KEYWORDS
SEGMENT
SOURCE i of 3
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 7258)
AUTHORS Gorlach,A., Lee,P.D., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
REFERENCE 2 (bases 1 to 7258)
AUTHORS Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Characterization of the genomic structure of the p47-phox gene
Unpublished
REFERENCE 3 (bases 1 to 7258)
AUTHORS Gorlach,A., Lee,P.D., Roesler,J., Chanock,S.J. and Curnutte,J.T.
TITLE Direct Submission
SUBMITTED (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source
location/Qualifiers
1. 7258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P40"
/chromosome="7"
/map="7q11.23"
/note="Human genomic p1 clone P40 is a p47-phox pseudogene
which is defined by the presence of a G deletion at the
beginning of exon 2"
<19..90
/gene="NCF1"
/pseudo
/number=1
3252..3330
/gene="NCF1"
```

```

/pseudo
/number=2
5084..5159
/gene="NCF1"
/pseudo
/number=3
5259..5424
/gene="NCF1"
/pseudo
/number=4
6783..6838
/gene="NCF1"
/pseudo
/number=5
BASE COUNT 1573 a 2009 c 1834 g 1842 t
ORIGIN
Query Match 38.2%; Score 1260.4; DB 9; Length 7258;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1275; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 2020 actggaggccaccagtcagtggggacaccttcacgtcgcacgtcgtgggttt 2079
Db 1 ACTGGAGGCCACCCAGTCATGGGGAGACACCTTCATCGGTCACATCGGCCTGCTGGGCTT 60
Qy 2080 gagaagccttcgtaccagccagcactatgtgagctggtgagggcat-cccgtgg 2138
Db 61 GAGAAGCGCTTCGTACCCAGCCAGCACTATGTGAGTAGTGGTGGAGGGCATCCCGTGG 120
Qy 2139 ggggaatacggaggagcagcacggccaccttgcagtcaccagggcccaaccagctccagt 2198
Db 121 GGGGAATACGGGAGGGAGCAGCACGCCACCTTCGAGTCCCGAGGCCAACCCAGCTCCAGT 180
Qy 2199 gaggactaaaggggcagggtcttggggacacctggtcccttggtcttttgagccttgatctacc 2258
Db 181 GAGACTAAACGGGCGAGGGTCTTGGGCACCTGTGGTCCCTGGTCTTTGAGCCTGAGTCTACC 240
Qy 2259 cctgtgacctcgtggaagacagttcccttgagaccgcctgagcccgccagcccttactgt 2318
Db 241 CCTCTGATCCCTGGGAGACAGTTCCTTGGACCCGCCCTGGGCCCCAGCCCTTTACTGT 300
Qy 2319 cccgcctctgtcccccagcaggccctcagcttagccagagtcctcttctctctctctccc 2378
Db 301 CCGCGCCTGTGTCCCGCAGCCAGGCCCTCAGCCTTAGCCAGGAGTCTCTTTCTGCTCCCC 360
Qy 2379 tgccatggcaggcagccagcgcctctctcaggtccagggcccaactctccagagaacct 2438
Db 361 TGCCATGGCCAGGCGCCAGCGCTCTCTCAGTCCGAGGCCCACTCTCTCCAGGAAGCCT 420
Qy 2439 tccctgactagccagcactatcagagagtgccctcccaagaggagcctggaactaaa 2498
Db 421 TCCCTGACTAGCCAGCTATCAGAGAGTGGGCCCTCCCAAGAGGAGGGCCTGGAACATAA 480
Qy 2499 gctctctctccccagctgcctgtagtgcagtagtagtcttctctctccagtagggt 2558
Db 481 GCTCTCTCTCCCCAGCTGCTGTAGTGCAGTGTAGAGTCTTATCTCTCTCCAGTAGGCT 540
Qy 2559 gacaccatgacaggggccaatagagtcctcccatctgtccccaaggaggctggacaaatg 2618
Db 541 GACACCATGACAGGGGGCAATAGAGTCTCTCCCATCTGTCCCAAGAGGGCTGAGCAATG 600
Qy 2619 cctgctcagacacacagtcacatgggtccctcccaatccataggaagccagggaggaac 2678
Db 601 CCTGCTCAGACACACAGTCCACCTGGGTCCCTTATCCATAGGAAGGCCAGGAGGAGAAC 660
Qy 2679 tacatttaggaattgaagcttgatggaacatttagtccttagtgccaagaccctttctc 2738
Db 661 TACATTTAGGAAATTGAAGCTTGTATGGAACTTTAGGCCCTATGTGCCAAGACCTTTCTC 720
Qy 2739 tttttgtttttttgtgttttgagacagagctctgatctgtgtgccagggccagagtg 2798
Db 721 TTTTGTGTATTTTGTGTTTTTTGAGACAGAGTCTTGATCTGTTGCCCGCCAGGAGATG 780
```


2811	Qy	ctcagctcactgcaaacctccgcctctccagggttcacactgggttctcctgcctccacgctccacg	2870
		CTCAGCTCACTGCAACCTCCGCCTTCCAGGTTCAACTGGTTCTCCTGCCCTCAGCCTCCAG	840
2871	Qy	agtagtgggattacagagtgcccaccaccacgcctggcctaatttttgatttttagtaga	2930
		AGTAGTTGGGATTACAGGTGCCACACACACGCCTGGCTAATTTTGTATTTTAGTAGA	900
2931	Qy	gcacagggttcaccatgttgccagagctggctctcaaaactcgaactcgaactcgaagtatccacc	2990
		GACAGGGTTTCACCATGTGGCCAGACGTGGTCTCAAACTCTGACCTCAAGTGATCCACC	960
2991	Qy	caactgggctcccaaatgtctgggattacaggcatagcacacgtgcctggcctgtttt	3050
		CACCTGGGCTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGTGCCCTGCTTTT	1020
3051	Qy	tttgaatgagtgctggagtgagtggtgcgatcatagtttaactgcaagctcaactccc	3110
		TTTGAATGAGTGCTGGAGTGAAGTGTGGCATCATAGTTCACTGCAAGCTCAACCTCCC	1080
3111	Qy	aggcccaagtatcctcctcctcctcaagcccttgagtgcctgagctggggctacaggcgcacaca	3170
		AGGCCCAAGTGATCTCTCTGCTCAGCCCTTGAGTAGCTGGGCTACAGGTCACACCAC	1140
3171	Qy	caatgcctgagctagtttttaaaattttgtggagataggtttcaactatgtctgcaggc	3230
		CAATGCTCGCTAGTTTTTAANAATTTTGTGGAGATGAGGTTTCACATATGTTGTCCAGGC	1200
3231	Qy	taacttggaactcctcctcggtctaagcaacccctctggtctcagcctcccaagtgctaggat	3290
		TAACTTGAACCTCTCGGTTTAAGCAACCCCTCTGGTCTCAGCCTCCACAGTCTAGGAT	1260
3291	Qy	tacaagcgtg	3300
1261	Db	TACAAGCTG	1270

Search completed: January 11, 2002, 16:31:36
Job time: 26648 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:23:10 ; Search time 814.02 Seconds
(without alignments)
3539.803 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaaggtcaggcactgtgc.....ctgtccaagtgggtgcatt 3361

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	406	12.1	52216	22	AAH28355
C 2	391.4	11.6	32152	22	AAI57791
C 3	385.8	11.5	51474	22	AAF97846
C 4	382.8	11.4	119950	20	AAAX90201
5	375.2	11.2	1624	22	AAH98793
6	373.2	11.1	72604	20	AAZ10752
C 7	370.6	11.0	32367	19	AAV35620
C 8	370	11.0	1298	22	AAF64178
C 9	370	11.0	162450	21	AAZ86967
C 10	369	11.0	15577	19	AAV35616
11	367.6	10.9	32192	22	AAI63523

C 12	366.6	10.9	106746	21	AAA10225
C 13	366	10.9	32134	22	AAI63522
C 14	363.6	10.8	160552	22	AAD02697
C 15	363	10.8	236303	22	AAI11614
C 16	362.6	10.8	8427	22	AAI57755
C 17	359.6	10.7	160552	22	AAD02697
C 18	359.2	10.7	23603	22	AAI62936
C 19	358.4	10.7	4352	22	AAI58667
C 20	358.4	10.7	4377	22	AAI60453
C 21	358.4	10.7	72604	20	AAZ10752
C 22	357.4	10.6	31853	22	AAI63343
C 23	355.2	10.6	1887	22	AAI61210
C 24	354.8	10.6	5301	22	AAF97866
C 25	354.2	10.5	24218	22	AAI62935
C 26	353	10.5	110000	22	AAF84800
C 27	352.2	10.5	122186	22	AAC89560
C 28	351.6	10.5	4124	22	AAI1154
C 29	351.2	10.4	122186	22	AAC89560
C 30	350.4	10.4	16891	20	AAI37084
C 31	350.4	10.4	161425	22	AAH02340
C 32	350.4	10.4	162025	22	AAH02339
C 33	349.8	10.4	65666	21	AAI53450
C 34	349	10.4	23613	22	AAI62930
C 35	348	10.4	32204	22	AAI57790
C 36	347.4	10.3	3884	22	AAH98627
C 37	346.2	10.3	986	21	AAH79767
C 38	345.6	10.3	52216	22	AAH28355
C 39	345	10.3	66685	22	AAI07380
C 40	344.8	10.3	2562	22	AAH17646
C 41	344.8	10.3	114793	22	AAD08215
C 42	344.6	10.3	32249	22	AAI62932
C 43	343.6	10.2	2483	22	AAH16966
C 44	343.4	10.2	14919	22	AAI63992
C 45	343.2	10.2	32190	22	AAI62927

ALIGNMENTS

RESULT 1
AAH28355/C
ID AAH28355 standard; DNA; 52216 BP.
XX
AC AAH28355;
XX
XX 05-SEP-2001 (first entry)
XX
DE Nucleotide sequence of the human musashi promoter.
XX
XX Musashi promoter; multipotential neural progenitor cell;
KW neural stem cell; central nervous system; ss.
XX
XX Homo sapiens.
XX
XX WO200146384-A2.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35395.
XX
PR 23-DEC-1999; 99US-0173003.
XX
XX (CORR) CORNELL RES FOUND INC.
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Goldman SA, Okano H;
XX
DR WPI; 2001-4i8053/44.
XX
PT Separating multipotential neural progenitor cells from a mixed
PT population of cells, involves introducing nucleic acid molecule
PT encoding fluorescent protein under promoter control, and separating
PT fluorescent cells -

XX Claim 12; Fig 11A-JJ; 87pp; English.

XX The present sequence represents the human musashi promoter, which is used in the method of the invention. The specification describes a method for separating multipotential neural progenitor cells from a mixed population of cell types. The method comprises introducing a nucleic acid encoding a fluorescent protein under control of a promoter which selectively functions in the progenitor cells, into all cell types of the mixed population, allowing only the progenitor cells to express the fluorescent protein, and identifying and separating cells that are fluorescent, where the separated cells are progenitor cells. The method is useful for isolation and purification of multipotential neural progenitor cells, especially neural stem cells from adult brain. The isolated cells are used in both basic analyses of precursor and stem cell growth control, as well as in more applied studies of their transplantability and engraftment characteristics. The cells are useful in support of the structural repair of the damaged central nervous system, such as in the traumatized brain, or the contoured, traumatized or transected spinal cord.

XX Sequence 52216 BP; 13494 A; 12461 C; 12362 G; 13899 T; 0 other;

Query Match 12.1%; Score 406; DB 22; Length 52216;
Best Local Similarity 54.3%; Pred. No. 3,4e-83;
Matches 1401; Conservative 0; Mismatches 1015; Indels 166; Gaps 22;

Qy 6 gtcaggcactgtggctcatgctgttaatccacgactttggaggccgagcggttgaat 65
Db 11688 GCCGGCAGCGGTCTCATGCTGTAATCCAGCACTTTGGGAGGCCGAGGGTGGAT 11629

Qy 66 cacttgagtgaggatctgagacagcctgacaataatcgtgaaactcc-atctctact 124
Db 11628 CAC--GAGTTCAGGAGATCGAGACCATCTGGCTAACATGGTGAACCCCTGCTCTACT 11571

Qy 125 aaaaatacaaaaattagccagggtgtgtggcggtgtgttagtcccgactacttgggag 184
Db 11570 AAAATATAAAAATTAGCTGGSCATGGTGTGTGCTATAGTCCAGCTACTTGGGAG 11511

Qy 185 gctgaggcaggtgaattacttgaacctggaactgtgaggtgtgcaatgagcaagattgca 244
Db 11510 GCTGAGGCAAGAGAATTGCTTGAACCCGGAGGAGAGAGGTTGTGTGAGCCAAAGATCATG 11451

Qy 245 ccactgcactcca-----gtgacagagcagactccactcaaaaaaataaaaaa 298
Db 11450 CCAGTGCACCTCCAGCCTGGGTGACAGAGCAAGACTTTGTCTCAAAAAAATAAAAAA 11391

Qy 299 agttgggaaagccagggtgagtgctgcacgcctgttaatcccaacactttaagagct 358
Db 11390 A---AAAAAAGGTCCTGTGTCGGTGCT--GAGCCCGTAATCCAGCACTTTGAGAGGCC 11335

Qy 359 gagggtggagaactcttgagccagaggttcgagacagcctgggaattgtcccaagac 418
Db 11334 AAGGTGGGTGGATCA--TGAGATCAGGAGTTCAAGACCAGCCTACCCAAATAGTGAAC 11277

Qy 419 ctgtcttttacaataaa-----ttagccgggtgtgtggcatacgtctgtgtccca 470
Db 11276 TCCGCTCTACTAAAAACACAAACTTAGTTGGGCATGGTGGCGTGTGCTGTAGTCCCA 11217

Qy 471 gctattcggagagctgagcagggagattgtttgagccttaggagctgagggctgtagtga 530
Db 11216 GCTGTCCAGAGGCTGAGGTAGAGATACTTGAACCCAGGAGGAGGTTGAGTGA 11157

Qy 531 gctgtgatacagctcaactgtacttgaactgagcagcagcagcagcagcagcagcagc 590
Db 11156 GCTAAGACTGCACCACTGCATCCAGCTGGGCAACAGGGCAAACTGTGCTCAAAAAA 11097

Qy 591 agaaaaataaagtgggaaggtcactcaactcactcagatgagacaagacatgtttga 650
Db 11096 AAATTGTATAATATAATAATCTATAATAATATATATATATAGAGAAACTAGAGGAATAA 11037

Qy 651 agtgtgagccgaagcctggagaacgctatgcgccaggaatgcagggcagcagagact 710

Db 11036 ATTAAATGTTAATA-----TGGTTGCCTCAATAATATGCGGATATATTT 10989

Qy 711 caagatgcagcgcctgttctggagccagatggccctgcaatgcccaactcaccctgc 770
Db 10988 TATTTAATCAACACATATTTCTTTAAGAAACTGTGTTTATTTAATGACATTTTGTGTTT 10929

Qy 771 cctccctctggccagacatcacccggcccccatactctcgagacgtaccgcgccttgc 830
Db 10928 TTTGGTTTTTTTTCAGATGGAGTCTTGTCTC-----TGTGCAAACTGGAGTGCAGTGGC 10873

Qy 831 cgactacgagaagacactcggtcctcgagatggctctgctccacgggggagcgtgtgagagt 890
Db 10872 GAGATCTCTGCTCACTCAACACTCCGCCTCCAAAAGTTCAAGCGATTTCTCTGCTCAGCC 10813

Qy 891 cgtgagaagcgcagcgcgtgcagacctcccaactcccaactccacgttccctcctccttgc 950
Db 10812 TCCTGAGTAG--CTGGGACTACAGGTGCGCCTCACCATGCCAGCTAATTTTTATATTTT 10754

Qy 951 caggaaaccacagccacaagccccctgccaaggtccagggcagcctggccccctggagac 1010
Db 10753 AGTAGAGACAGGG-----TTTCACCATGTTGGATACAGGTG 10717

Qy 1011 tccagctgttaggggcccctaaatgtcctccccacactgtgggtcgcctctctctcttag 1070
Db 10716 TGAGCCTCCGTCCCGCGCTGTATTTTCTTTAATAAATATATATAGAATATATATAGC 10657

Qy 1071 tgtgcaccctgt 1130
Db 10656 AAAACACATTTCTTTTGTGTTTGTGTTTGTGAGACAGAGTCTTGTCTGTCAACCCAG 10597

Qy 1131 cgtgtctgtctgt 1190
Db 10596 GCTGTATAG-----CAGTGGCATGATCTCTGCTCAGTGC 10563

Qy 1191 gattctgtgacatgtgtgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 1250
Db 10562 AACCTCTGCTCCGGTTCAACAATTTCTCTGCTCAGCCTCCCAAGCAGCTGGGATG 10503

Qy 1251 acatttccaaatctgacatgtgacatcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgt 1310
Db 10502 ACAGCGCGCTGCCACCCCGCGCTAATTTTGTATTTTAGTAGAGATAGAGTTTTCAC 10443

Qy 1311 ctgagtgtgtgagggagcgcgtggccctgcctccagtcacattcccgcaactctgtgca 1370
Db 10442 CATGTTGGCCAGGCTGTGCTGAACCTCTGACCTTAGTGTATCCACCCACCTCAGCCTCC 10383

Qy 1371 caggt 1430
Db 10382 CAATGTGTAG-----GATTACAGCATGAGCCACTGCATCCAGCCCTTAACACCAT 10331

Qy 1431 agccccctgacagtcctgacgagacggaagaccctgagcccaactatcaggtgtccccct 1490
Db 10330 TTCATATTAGTAGTTTTTTA-----AAAAATTACAAATGTAAACAT 10293

Qy 1491 gccctcccgaggctgtagggtgtgtggagaaaggcagggcagggcagggcagggatattgagt 1550
Db 10292 GTGCTCACA-----TTAAAAAGAAAAAATCGCTCAGTACACAAAGGTATAAGATG 10243

Qy 1551 gactccttggagctgtggctgt 1610
Db 10242 AAAAAAGTAAAGTTCTTTACCCCTTAGACTCTCTCCCTTAAATGTAACTGTACCAATTTTC 10183

Qy 1611 tgggtctgt 1668
Db 10182 ATGTATATCTTCCAAAAAATGTTCCTATATACCTATATACCTATATATATATATATTTTA 10133

Qy 1669 agggagggagcagacgactcgggagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1728
Db 10122 AAAAAATAGATATAGCTACTTGGGAGGCTGAGTGGGAGGATCACTTGGAGCCCAAGAGG- 10062

Qy 1729 caggtctcagtggaagctgaaggtgagcagaccttagctcaggaagagggcctgtgtgtgt 1788

[illegible]

QY 688 ggaatgacagggcagagactcaagatgccagcgc-----ctgtcttgaggcccca 740
Db 29504 CACGAATCCCTACTGTGTAGACATTTGATTGTCACATTTATCATCTGTATACATGGAGC 29445
QY 741 gatggccctcaatgcccactaccctgcctcctcctctgtcccccagacatcacggccc 800
Db 29444 GATGCTCCACGCTTGAGTTTAAAGAAATTCAGACCTGCCAGGAGCCCTGTGTAAAGAGCC 29385
QY 801 catcatctgcagactgaccgcgccattgcccactacagagaactcctgggctcccgaga - 859
Db 29384 TAGGGAGGCCAGAGCCGCTGCCCTACGCCCTGTATCTTAACACTTTGGGAGCCGAGAC 29325
QY 860 ----tggctctgtccacggggagcgtggtgaggtcgtgagagaagcagagcggctcag 915
Db 29324 AGGCAGATTGCTGAGCTCAGGAGTTTGAGACACAGCCAGGCGAACACCGGAAACCCCGT 29265
QY 916 acctcccacattacgggctccttccctcgtgtcagaagaacccacagcccaagccccc 975
Db 29264 CTCCTACTAAATACAAAAAGTAGCCGGCGCTGGTGGGCACCTGTAGTCCCAAGCTAC 29205
QY 976 tgcgaagctcagggcagcctggccctgggaggaactccagctctgttagggccctaaat 1035
Db 29204 TCGGAGGCTGAGGAGAGAAATTCCTTGAAACCGGAGGCGAGGTTGCAGTGATCCAA 29145
QY 1036 gtccctcccacactggtggtgcgctctctctcttagtgcaccctgtggtgctggtg 1095
Db 29144 GATCGCAC -CACTGCACCTCCAGCCTGGGTGACAGCGAGACTCCGCTCCCAAAAAA 29087
QY 1096 atctgtgatggcagcgggggggcatgtctcgtgttctgtctgctggatgggtatggg 1155
Db 29086 AAAAAAGAGCCAGGCGAGACTGCAAGAAGCCACATGCAAGGTGAGGAAGAGTAGGC 29027
QY 1156 accgtctgttattatgaagtgggtcagagctgtgattctgtgag-----ca 1203
Db 29026 CAGCTCAGGTATCTAGGAAACTCTTAGCTGCCATGCAACAGGTCAGTAGTGCCA 28967
QY 1204 tgtgtgatgatgatgtgacctatgtccagttgtcagttgtgtgaggtgacattccaaatc 1263
Db 28966 TCAGAGTTAGGAGAGGTTGGGCTCAGGAGAAATTTCTCCTCAGTTTGTCTCCAGGTGCAT 28907
QY 1264 tgagcattggacatcagttgtctgtgtcctctgtgtcctcaccatccctgatggctgag 1323
Db 28906 CCCCCTGTAGACTTAAGCCCCCACAGAGAGATGGGTAGCATAAAGTGAAGAACCACAA 28847
QY 1324 gtagcgtctggccctgccctcagtcacattccgcacacctctgcacaggtgtgtgtt 1383
Db 28846 GTCCGGGATTTGGCTGTGCTGTAGTCTCAGTCTGTTGGGAGGTTGGGTTGGGAGGAT 28787
QY 1384 c---tgtcagatgaagcaaaagcaggtgtggtatcccgacatcttctcgcagccctgga 1440
Db 28786 GGCCTTGAGCCAGGAGTTGTGTCTGCCCTGGGCAACACAGCTTTATTAGGAAAAAGGA 28727
QY 1441 cagtctcagcagacggaagaccctgagcccaactatgcaggtgccccctgcctcccgag 1500
Db 28726 AGGGAAGGGGAAGGGAAGGAGAGAGAGGAAGGAAGGAGGAAGGAGACAGAGAGAAGG 28667
QY 1501 gctgtagggtgtggagaaagggcaggcagg-----ctca 1538
Db 28666 ACAGAGCTGAGAGAAAGAAAGGAAGGAGGAAGGAAGGAAGGAAGGAAGGNA 28607
QY 1539 gggatattgagtactgtttggagctctgggctgtgtgtgtgtgtggcagaaaaagtcagg 1598
Db 28606 GGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 28547
QY 1599 gctaagatctcaggtctgt 1658
Db 28546 GGAAGGAAGGAAGGAAAGGAAAGGAAGGAAAGGAAAGGAAGGAAGGAAGGAGAG 28487
QY 1659 acaggaaccagagagagagcagcagctcggagaggtggagagccagtgtgtgtgtgt 1718
Db 28486 AGAGAAAGAAAGAAAGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 28427
QY 1719 gatatgtggccaggttctcagttggaagctgaaggaagcagacaccttagctcaggaagga 1778

Db 28426 GAGGAGGGAAGGAAGGAAGGAGGGAAGGAGGGAAGGAAGGAGGAAGGAAGGGA 28367
QY 1779 gggctgctggaagtgggggcatcatcactaccagaaggaagaaactggcagtg----- 1833
Db 28366 GGGAGGAGGGAAGGCGGAGAGAGAGAGAGAGAAAGAAAGAAAGAAAGAGAGGA 28307
QY 1834 -----ccagggtcgtggtgggctgcattgagcttgaaaaaacataataga 1881
Db 28306 GGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG 28247
QY 1882 atggttaccattttattttattttattttattttattttattttattttattttatttt 1941
Db 28246 TTTGGTTTGGGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 28187
QY 1942 cactcctctgaagctggagtcggtggtgctctcagctcactcgaactcgaactcgtcct 2001
Db 28186 TCCTGTGCACCCAGCGCTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTG 28127
QY 2002 ccaggaLcaagtgattctccagctcagctcccccagctagctgggattacaaagcatgc 2061
Db 28126 CCCAGGTTCAAGCAATTCCTGTCTCAGCCCTCCCA -GTAAC TGGGACTACAGGTTAC 28068
QY 2062 accacatgctcgggataatttttattttattttattttattttattttattttatttt 2121
Db 28067 GCCACAGGCCCGAGCTAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTT 28008
QY 2122 agactggtctgaactctgacactcagctgagctcctcagctcagctcagctcagctcag 2181
Db 28007 AGGCTGGTCTCGAATTCCTGACCCGAGGTGATCACCACCTCGGCTCCCAAGTGCTG 27948
QY 2182 gaattacagatgtgagcactgctcctggcctggttacccacatttttaaaatggagtgat 2241
Db 27947 GGATTACAGGCATGAGCCACTGCTGCTGGC----- 27916
QY 2242 ttacccctttatgtggatttatcagctgtgttttttttttttttttttttttttttttt 2301
Db 27917 -----CATTTT----- 27880
QY 2302 tctgtcacccagctggagtcagtaataatcagctcagctcagctcagctcagctcagctc 2361
Db 27879 TCTGTACCCAGGCTGGGGTGCAATGGCAGCATCTCAGCTCACACACCTCTGCCTCCC 27820
QY 2362 gggttcaagaattctcctcagctcagccactgagtagcctggggttacaggaatgcac 2421
Db 27819 AGGTTCAAGCAATTCCTGCTCAGCTCAGCTCTCAAGTAA -CTGGGACTACAGGCGCATGCC 27761
QY 2422 accagccaggtctaatt 2481
Db 27760 ACCAGCCTTAGCTAA -TTTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTT 27702
QY 2482 gctgtctcagactcagctcaggtgagtcgcccgccttgccctcccaagtgctag 2541
Db 27701 GTTGTCTTTGAATCCTTGACCTCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 27642
QY 2542 attacaggtgggaacacacttgcagcctgcccagcctgtggctatcgttt 2584
Db 27641 ATTACAGGCATGAACCACTGCGCGCGCCTTTT----- 27599
RESULT 3
AAF97846/c
ID AAF97846 standard; DNA; 51474 BP.
XX
AC AAF97846;
XX
DT 31-MAY-2001 (first entry)
XX
DE Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO:60.
XX
KW Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;
KW tumour suppressor; human lp36 homozygosity deletion domain; tumour;
diagnosis; ds.

02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488784/53.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
PS Disclosure; SEQ ID NO 837; 564pp + Sequence Listing; English.

Db 17653 actctgacctcagatgatccaccgccttgctgcctccgaagtctgggattataggcgt 17712

Qy 2553 gaaccacttgcgcagcctgtgctatcgttttaaac 2588
 || ||||| || | | | | | | | |

Db 17713 gagccaccactcctgctcaactcctctctctgtac 17748
 :
 :

RESULT 14

AA02697
 ID AAD02697 standard; DNA: 160552 BP.

XX
 AC AAD02697;

XX
 DT 02-MAY-2001 (first entry)

XX
 DE Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.

XX
 KW Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ds.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers

FT FT exon 32847..32922
 FT FT /*tag= a
 FT FT /number= 1
 FT FT /label= 4a_5U4
 FT FT 32923..35592
 FT FT /*tag= b
 FT FT /cons_splice= (5'site:NO, 3'site:YES)
 FT FT 35593..35674
 FT FT /*tag= c
 FT FT /number= 2
 FT FT /label= 4a_5U3
 FT FT 35675..45093
 FT FT /*tag= d
 FT FT 45094..45185
 FT FT /*tag= e
 FT FT /number= 3
 FT FT /label= 4a_5U2
 FT FT 45186..46633
 FT FT /*tag= f
 FT FT /cons_splice= (5'site:NO, 3'site:NO)
 FT FT 46634..46700
 FT FT /*tag= g
 FT FT /number= 4
 FT FT /label= 4a_5U1
 FT FT 46701..47938
 FT FT /*tag= h
 FT FT /cons_splice= (5'site:YES, 3'site:NO)
 FT FT 47939..49746
 FT FT /*tag= i
 FT FT /number= 5
 FT FT /note= "Includes 17 base pairs of 5'UTR, the ORF
 FT FT and all of 3'UTR"
 FT FT 47939..47955
 FT FT /*tag= j
 FT FT /note= "Portion of 5' untranslated region (5'UTR)"
 FT FT 47956..49128
 FT FT /*tag= k
 FT FT /product= "Human glycosyl transferase-4alpha
 FT FT (GST-4alpha)"
 FT FT 49129..49746
 FT FT /*tag= l
 FT FT 83257..83347

FT FT /*tag= m
 FT FT /label= 4a_5U2
 FT FT 83348..96412
 FT FT /*tag= n
 FT FT /cons_splice= (5'site:NO, 3'site:NO)
 FT FT 96413..96484
 FT FT /*tag= o
 FT FT /label= 4a_5U1
 FT FT 96485..98456
 FT FT /*tag= p
 FT FT /cons_splice= (5'site:NO, 3'site:NO)
 FT FT 98457..99968
 FT FT /*tag= q
 FT FT /note= "Includes 17 base pairs of 5'UTR, the ORF
 FT FT and all of 3'UTR"
 FT FT 98457..98473
 FT FT /*tag= r
 FT FT /note= "Portion of 5' untranslated region (5'UTR)"
 FT FT 98474..99661
 FT FT /*tag= s
 FT FT /product= "Human glycosyl transferase-4beta
 FT FT (GST-4beta)"
 FT FT 99662..99968
 FT FT /*tag= t
 XX
 PN W0200106015-A1.
 XX
 XX 25-JAN-2001.
 PD
 XX
 PF 19-JUL-2000; 2000WO-US19741.
 XX
 XX 20-JUL-1999; 99US-0144694.
 PR
 PR 13-JUL-2000; 2000US-0593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 PI Rosen SD, Lee JK, Hemmerich S;
 XX
 XX WPI: 2001-138471/14.
 DR
 DR P-PSDB; AAY72639, AAY72640.
 XX
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications -
 XX
 PS Example 1; Page 62-104; 128pp; English.
 XX
 CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
 CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
 CC chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand. GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX
 SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 10.8%; Score 363.6; DB 22; Length 160552;
 Best Local Similarity 53.9%; Pred. No. 2.8e-73;
 Matches 1401; Conservative 2; Mismatches 991; Indels 204; Gaps 24;
 Qy 2 aaaggtcaggcactgtgctcgtcgtatccacccacacttggaggccgagcggtt 61


```
QY 2160 gctcggcctcccaagtgctggaattacagatgtgagccactgtccctggcctggttac 2219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140311 cctcagcctcctaagtgctgggattacagcgtagccactacgctggcctgagttg 140370
QY 2220 ccacattttaaaatggagtgattccacccttttatgtgg- ---attcacagctgtgtttt 2275
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140371 tgcacttaaatgggtacattttatggtgtggaattgtgtctcaataaggctttttttt 140430
QY 2276 ttttttttttgagacaaagtctggctctgtccaccaggctggagtgcaagtaatgcaatc 2335
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140431 ttttttttttgagcggagtgctcctctgtccaccaggcgagtgcaagtgccgcatc 140490
QY 2336 tcaagtcactgcaaccttagcctctctggtttcaagaattctcctgcagccaccctga 2395
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140491 tcgggtcactgcaagctcgcctccacagttcattgccattctccggctcagcctcccca 140550
QY 2396 gtacgtcgtgggttacagcgtatgcaccacacgcccaggtcaatttttttta 2454
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140551 gtac-ctgggactacagcgctgcaccacacgcccggctcaatttttatatttttagta 140609
QY 2455 gagatggggtttcgcatgttgccaggctgtctcgaactctgacctcaggtgatccg 2514
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140610 cagacggggtttcaccggttagccaggatggtctcgatctcctgacct--tgtgatccg 140667
QY 2515 cccgccttggtccctccaa 2532
||| ||||| ||||| |||||
Db 140668 cctgcctcagcctcccaa 140685

RESULT 15
AAS11614/c
ID AAS11614 standard; DNA; 236303 BP.
XX
AC AAS11614;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human genomic DNA containing exons 2-17 of the CRIM1 gene.
XX
KW CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
neuroprotective; renal; osteopathic; dental; vulnerable; immunogen;
antibody; gene therapy; neurodegenerative disease; eye disorder;
KW cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
tooth abnormality; wound; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT exon 33104..33277
FT /*tag= a
FT /number= 2
FT intron 33278..77746
FT /*tag= b
FT /number= 2
FT exon 77747..77989
FT /*tag= c
FT /number= 3
FT intron 77990..79103
FT /*tag= d
FT /number= 3
FT exon 79104..79224
FT /*tag= e
FT /number= 4
FT intron 79225..101022
FT /*tag= f
FT /number= 4
FT exon 101023..101144
FT /*tag= g
FT /number= 5
FT intron 101145..113377
FT /*tag= h
FT /number= 5
```

```
FT exon 113378..113560
FT /*tag= i
FT /number= 6
FT intron 113561..115985
FT /*tag= j
FT /number= 6
FT exon 115986..116183
FT /*tag= k
FT /number= 7
FT intron 116184..135707
FT /*tag= l
FT /number= 7
FT exon 135708..135836
FT /*tag= m
FT /number= 8
FT intron 135837..146471
FT /*tag= n
FT /number= 8
FT exon 146472..146628
FT /*tag= o
FT /number= 9
FT intron 146629..148761
FT /*tag= p
FT /number= 9
FT exon 148762..148883
FT /*tag= q
FT /number= 10
FT intron 148884..150044
FT /*tag= r
FT /number= 10
FT exon 150045..150254
FT /*tag= s
FT /number= 11
FT intron 150255..153815
FT /*tag= t
FT /number= 11
FT exon 153816..154031
FT /*tag= u
FT /number= 12
FT intron 154032..158580
FT /*tag= v
FT /number= 12
FT exon 158581..158802
FT /*tag= w
FT /number= 13
FT intron 158803..173982
FT /*tag= x
FT /number= 13
FT exon 173983..174177
FT /*tag= y
FT /number= 14
FT intron 174178..181006
FT /*tag= z
FT /number= 14
FT exon 181007..181129
FT /*tag= aa
FT /number= 15
FT intron 181130..183612
FT /*tag= ab
FT /number= 15
FT exon 183613..183800
FT /*tag= ac
FT /number= 16
FT intron 183801..185152
FT /*tag= ad
FT /number= 16
FT exon 185153..187765
FT /*tag= ae
FT /number= 17
XX
PN WO200138519-A1.
XX
PD 31-MAY-2001.
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 16:31:36 ; Search time 14958.6 Seconds
(without alignments)
3706.691 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaagtcagcagctgtggc.....ctgtccaagtgggtgcatt 3361

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo_hum.*
- 31: em.htgo_inv.*
- 32: em.htgo_rod.*
- 33: em.htg_hum.*
- 34: em.htg_inv.*
- 35: em.htg_rod.*
- 36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3312.8	98.6	124526	9	AC005080	Homo sapi
2	3306.6	98.4	230552	9	AC005098	Homo sapi
3	3304.2	98.3	131359	9	AC004883	Homo sapi
4	3299.4	98.2	169604	9	AC003884	Homo sapi
5	3293.2	98.0	275197	9	AC004166	Homo sapi
6	3254.2	96.8	17302	9	AF184614	Homo sapi
7	2993.2	89.1	3348	9	HS47P43S02	Homo sapi
8	2990	89.0	3350	9	HS47P43S02	Homo sapi
9	2984.4	88.8	3349	9	HS47P43S02	Homo sapi
10	2843.8	84.7	3200	9	HS47P43S02	Homo sapi
11	2823.6	84.0	2899	9	HS47P43S02	Homo sapi
12	2668.8	79.4	184558	2	AC092405	Papio cyn
13	1719.8	51.2	149830	2	AC027219	Homo sapi
14	886.4	26.4	918	9	HSNCF1S4	Homo sapi
15	870.8	25.9	904	9	HS39NCF5	Homo sapi
16	772	23.0	799	9	HS39NCF6	Homo sapi
17	505.8	15.0	124095	9	AC002477	Human PAC
18	500.8	14.9	149830	2	AC027219	Homo sapi
19	497.2	14.8	149490	2	AL589985	Homo sapi
20	489.6	14.6	69901	9	HS694E4	Human DNA
21	482.6	14.4	315835	2	AL591212	Homo sapi
22	481	14.3	178451	9	AL139396	Human DNA
23	476.8	14.2	188833	9	HS268H5	Human DNA
24	472.4	14.1	178451	9	AL139396	Human DNA
25	464.8	13.8	191717	2	AC012482	Homo sapi
26	464.4	13.8	180675	2	AL354671	Homo sapi
27	464.4	13.8	240501	2	AL590226	Homo sapi
28	459.6	13.7	123829	9	HSJ469A13	Human DNA
29	459.6	13.7	181386	9	AC007686	Homo sapi
30	458.6	13.6	115079	2	AC008105	Homo sapi
31	457.8	13.6	107808	2	AL162271	Homo sapi
32	457.8	13.6	213432	2	AC068198	Homo sapi
33	456.4	13.6	145833	2	AC026539	Homo sapi
34	456.2	13.6	163035	2	AC006405	Homo sapi
35	456.2	13.6	180407	2	AL590822	Homo sapi
36	455	13.5	166549	2	AC025990	Homo sapi
37	455	13.5	183996	9	AF196779	Homo sapi
38	454.8	13.5	240455	2	AC010290	Homo sapi
39	454	13.5	120955	9	HUAC002310	Human Chr
40	452.8	13.5	167515	2	AC091864	Homo sapi
41	452.2	13.5	120955	9	HUAC002310	Human Chr
42	452.2	13.5	162365	2	AC018521	Homo sapi
43	448.6	13.3	180236	2	AC020600	Homo sapi
44	448.2	13.3	208774	2	AC009634	Homo sapi
45	447.8	13.3	99876	9	HSDJ90108	Human DNA

ALIGNMENTS

RESULT	1					
AC005080						
LOCUS	AC005080	124526 bp	DNA	PRI	02-OCT-2000	
DEFINITION	Homo sapiens BAC clone CTA-269p13 from 7q11.2, complete sequence.					
ACCESSION	AC005080					
VERSION	AC005080.2	GI:7770715				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	1 (bases 1 to 124526)					
JOURNAL	Sulston,J.E. and Waterston,R.					
MEDLINE	Toward a complete human genome sequence					
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)					
AUTHORS	99063792					
TITLE	2 (bases 1 to 124526)					
JOURNAL	Scott,K., Layman,D., Kalicki,J. and Harmon,G.					
	The sequence of Homo sapiens BAC clone CTA-269p13					
	Unpublished					

```

REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 124526)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 124526)
Waterston,R.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 12, 2000 this sequence version replaced gi:3212911.
-----
Center: Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_RG269P13
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-269P13 is from a release of the human BAC library CTRB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-754G14; the clone sequenced to the right is RP11-396K3. Actual start of this clone is at base position 1 of CTA-269P13; actual end is at base position 124526 of CTA-269P13.

The clone CTA-269P13 may contain a transposon in the growth of the clone, which is not part of the submitted sequence.

FEATURES

source

```

Location/Qualifiers
1..124526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-269P13"
/clone_lib="CTR-HS-A"
1..138
/note="match to EST AA643257 (NID:g2568475) nr61h09.s1"
1..338
/note="match to EST D44965 (NID:gl572440)"
42..118
/note="similar to EST D44969 (NID:gl572444)"
complement(106..396)
/db_xref="GI:3800554"
119..483
/note="similar to EST AW236834 (NID:g6569223) xm49f08.x1"
119..483
/note="similar to EST AI609282 (NID:g4618449) tw83h06.x1"
119..378
/note="similar to EST D44969 (NID:gl572444)"
146..483
/note="similar to EST N80671 (NID:gl243372) zb03a07.s1"
164..483
/note="similar to EST AW204636 (NID:g6504108)"
204..206
/note="match to EST AA643257 (NID:g2568475) nr61h09.s1"
306..652
/note="match to EST AA804480 (NID:g2875993) nz33c09.s1"
335..520
/note="similar to EST AW177547 (NID:g6443584)"
577..666
/note="similar to EST AW177547 (NID:g6443584)"
595..711
/rpt_family="Alu"
712..740
/rpt_family="(CAAAA)n"
741..898
/rpt_family="Alu"
899..919
/rpt_family="AT-rich"
965..1106
/rpt_family="Alu"
1107..1395
/rpt_family="Alu"
1396..1570
/rpt_family="Alu"
1571..1662
/rpt_family="L1"
1665..1817
/rpt_family="L1"
1834..1886
/rpt_family="L2"
1888..2189
/rpt_family="Alu"
2256..2546
/rpt_family="Alu"
3070..3391
/rpt_family="Alu"
3493..3780
/rpt_family="Alu"
3792..4101
/rpt_family="Alu"
4108..4232
/rpt_family="Alu"
4247..4422
/rpt_family="Alu"
4423..4729
/rpt_family="Alu"
4730..4866
/rpt_family="Alu"
4941..5099

```

misc_feature

misc_feature

misc_feature

STS

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region /rpt_family="L1"
5222..5527 /rpt_family="Alu"
repeat_region 5812..5883 /rpt_family="MERL_type"
repeat_region 5884..6194 /rpt_family="Alu"
repeat_region 6207..6397 /rpt_family="Alu"
repeat_region 6398..6469 /rpt_family="MaLR"
repeat_region 6470..6771 /rpt_family="MaLR"
repeat_region /rpt_family="Alu"
6772..7113 /rpt_family="MaLR"
repeat_region 7114..7422 /rpt_family="Alu"
repeat_region 7423..7468 /rpt_family="MaLR"
repeat_region 7469..7644 /rpt_family="MaLR"
repeat_region 7645..7774 /rpt_family="MERL_type"
repeat_region 7860..8155 /rpt_family="Alu"
repeat_region 8194..8466 /rpt_family="Alu"
repeat_region 8479..8599 /rpt_family="MIR"
misc_feature 8599..8600 /note="similar to EST M78139 (NID:g273876)"
repeat_region 8601..8916 /rpt_family="Alu"
misc_feature 8787..8803 /note="similar to EST A463503 (NID:g2188387) zx72cl0.s1"
misc_feature 8810..9112

Query Match 98.6%; Score 3312.8; DB 9; Length 124526;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3351; Conservative 0; Mismatches 7; Indels 4; Gaps 3;

Qy 1 aaaaagctcaggcaactgtggtcattcctgttaataccagcaactttggaggccgagcggt 60
Db 116133 AAAAGGTCAGGCACATGTGGCTCATCCCTGTAATCCAGCACATTTGGAGGCGGAGCGGT 116192

Qy 61 tgaatcaactaggtcagaggttcagaccagctgaccaatatcgtaaaaactccatctg 120
Db 116193 TGAATCACCTGAGGTCAGAGGTTCCAGACCAGCCCTGACCAATATCGTAAATCCATCTC 116252

Qy 121 tactaaaaatacaaaaattagcaggtgtggtggcgggtgcttgttagtcccagctacttg 180
Db 116253 TACTAAAAATACAAAATTAGCCAGGTGTGGTGCGAGGTGTGTAGTCCCAAGCTACTTG 116312

Qy 181 ggagctcaggcaggtgaattacttgaaacctggagggtggaaggttgcaatgagccaaagt 240
Db 116313 GGAGGCTCAGGCAGGTGAATTAATTAACTTGGAGGTGGAGGTTGCAATGAGCCAAAGAT 116372

Qy 241 tgcacactgcactccagtgcagagcgagactccatctc-aaaaaaaaaaaaaaaaa 299
Db 116373 TGCACACTGCACTCCAGTGACAGAGCGAGACTCCATCTCAAAAAAAAAAAAAAAAAA 116432

Qy 300 gttggaaaagccagggtgcagtggtccacgcctcgtaatcccaacacttttaaggaggtg 359
Db 116433 GTTGGGAAAAGCCAGGTGCAGTGGCTCCACGCCTGTAATCCCAACACTTTTAAGAGGCTG 116492

Qy 360 aggtgggaaatccctttgagccaggagttcgagaccagcctgggcatgtgtcccaagacc 419
Db 116493 AGGTGGGAAATCCCTTTTCAGCCAGGAGTTCCAGACCAGCCCTGGGCATTTGTCCCAAGACC 116552

Qy 420 ttgtctttcaaaaaattagcgggtgtgtgtgcaatagctgtgtgtcccaactattcgg 479
Db 116553 TTGTCTTTACAAAAAATTAGCGGGGTGTGGTGGCATACGTCTGTGTGCCCACTATTCCG 116612
```

```
Qy 480 gaggtcaggcaggagagattgcttgagcctaggagcttaggagcttagtgagctgtgatc 539
Db 116613 GAGGCTGAGGCGAGGAGATGTCTTGAGCTTAGGCTTAGGCTTAGGCTTAGTGTAGCTGTGATC 116672

Qy 540 acgtcactgtactctagctcgtgggcaacagagcaagactctgtctccaaaaaagaaaaaa 599
Db 116673 AGTCTACTGTACTCTTAGGCTGGGCAACAGACAGACTCTGTCTCCAAAAAAGAAATAA 116732

Qy 600 agttggaaaagctcactaaacttcatcagatagagcaaaaagacatgtttgaaagtgtgag 659
Db 116733 AGTTGGGAAAAGGCTCACTAACCTTCATCAGATCAGAACAAAAGACATGTTTGAAGTGTGAGG 116792

Qy 660 ccgaagcctggagaaacgctatgccccaggagaaatcagagggcgagagactcaaatgccc 719
Db 116793 CCGAAGCCTGGAGAACGCTATATCGGCCAGGAAATCAGGGCGAGCAGAGACTCAAGATGCC 116852

Qy 720 aqgcctctgttctggagggcccaagatggcctgaatgccactcaccctcgcctccctct 779
Db 116853 AGCGCTGTGTTCTGGAGGCCAGATGGGCCCTGCAATGCCCACTCACCTTGCCCTCCCTCT 116912

Qy 780 tgccccagacataccggcccccatcatctctgcagacgtaccgcgcattgcccgaactaga 839
Db 116913 TGCCCCAGACATCACGGGCCCATCATCTCTGCAGAGCTACCGCGCATTTGCCCACTACGA 116972

Qy 840 gaagacctcgggctccgagatggtctgttccacgggggagcgtggggaggttcgttgagaa 899
Db 116973 GAAGACCTCGGGCTCCGAGATGGCTGTGTCCACGGGGAGCTGGTGGAGGTGCTGGAGAA 117032

Qy 900 gagcagagcgttcagacctcccacattacgggggtcctctcccctgggtcgtcaggaacc 959
Db 117033 GAGCGAGAGCGGTTCAGACCTCCACCTTACGGGGGTCTTCCCTCTGGTCTCAGAGAACCC 117092

Qy 960 acagccacaagccccctgccaaggtcagagcagcctggccccctggagagactcccagctct 1019
Db 117093 ACAGCCACAAGCCCCCTGCCAAGGCTCAGGCAGCCTTGCCCCCTGGGAGGACTCCGCTCT 117152

Qy 1020 gttagggccctaaatgtctctcccaactgtggtgcgctctctctcttagtgtgaccc 1079
Db 117153 GTTAGGGGCCCTAAATGTCTCTCCACACTGTGTGGGTGGCTTCTCTCTTAGTGTGCACCC 117212

Qy 1080 tttgtgtgtgtggcactctgtgcatggcagccggggggggggaatgtctgcgtgtcttg 1139
Db 117213 TGTGTGTGTGTGGGCACTGTGTGCATGGCAGCCGGGGGGGGGCGCATGTCTGGGTGTCTG 117272

Qy 1140 tctgatatgggtatgggaccgtctgttcatatgaagtggtgcagagctgtgattctgtg 1199
Db 117273 TCTGTATGGGTATGGGACCGCTGTGTTCATTATGAAGTGGGCTCAGAGCTGTGATCTGTG 117332

Qy 1200 agcatgtgtcagtcgcatgcatgtgaacctcattgtccagtggtggtgaaggtgacattcca 1259
Db 117333 AGCATGTGTGCATGCATGCATGTGACCTCATTTGTCCAGTGTGGTGAAGGTGACATTTCCA 117392

Qy 1260 aatctgagcatgtgacatacagtgctgtgtctctgtctcctcaccatccctgatggct 1319
Db 117393 AATCTGAGCATTTGGACATCAGTGTGTGTGTCTGTCTCTGTCTGTCTGTCTGTGTGTGTGT 117452

Qy 1320 gcaggagcgctggccctgcctcagtcacatcccgacacatcccgacaccttggtgacaggtt 1379
Db 117453 GCAGGAGCGCTGGGCCCTGGCCCTCAGTACATTTCCCGCAGCTCTGCACAGGTTGCT 117512

Qy 1380 ggtctgtcagatgaagcaagcaggcgtggatccccagcatccttctctcagagccctgg 1439
Db 117513 GGTCTGTGCATGAAGCAAAAGCAGGCTGGATCCAGCATCTTCTCCAGCCCTCTGG 117572

Qy 1440 acagtcctgacagagcaggaagacctgagcccaactatgacaggtgccctcctccaga 1499
Db 117573 ACAGTCTGTACAGAGCGGAAGACCTTGACCCCACTATCAGGTGGCCCTGCCCTCCGA 117632

Qy 1500 gctctaggggtgtgggagaaagggcagcggttcagggatatctgagtgactgctct 1559
Db 117633 GCTCTGTAGGGGTGTGGGAGAAAAGGCGCAGGCGCTCAGGATATTGAGTACTGCTTT 117692

Qy 1560 ggaagtctggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1619
```


TITLE
JOURNAL
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
4 (bases 1 to 230552)
Waterston,R.H.
TITLE
JOURNAL
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
5 (bases 1 to 230552)
Waterston,R.
TITLE
JOURNAL
Direct Submission
Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:3212893.
COMMENT

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu

Summary Statistics

Center project name: H_RG350L10

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequencing of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send
mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-350L10 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-350L10;
actual end is at base position 230552 of CTA-350L10.

The sequence CTA-350L10 from base position 222330 to 222775 is a CA
rich region. The sequence is not exact but it is believed to be
the best representation of this region. The region was sized by
PCR from clone DNA at 650 bp. The region corresponds to
restriction digest hindiii: band size 7685 in silico and 7756 real.

FEATURES source

1. 230552
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.2"
/clone="CTA-350L10"

/clone_lib="CITB-HS-A"	repeat_region
441..725	
/rpt_family="Alu"	misc_feature
783..1149	
/note="similar to EST BE000775 (NID:g8261008)"	misc_feature
910..1312	
/note="similar to EST R22057 (NID:g776838) yh25f08.sl"	misc_feature
1286..1716	
/note="similar to EST H72453 (NID:g1044269) ys04f07.rl"	repeat_region
1609..1864	
/rpt_family="Alu"	misc_feature
1984..2271	
/note="similar to EST R22113 (NID:g776894) yh25f08.rl"	misc_feature
2337..2682	
/note="similar to EST AW383168 (NID:g6887827)"	misc_feature
2437..2558	
/note="similar to EST N56845 (NID:g1200735) yw83e04.rl"	repeat_region
2532..2601	
/rpt_family="MIR"	repeat_region
3054..3367	
/rpt_family="Alu"	misc_feature
3240..3391	
/note="similar to EST AI566941 (NID:g4525393) tn25b07.xl"	misc_feature
3588..3845	
/note="similar to EST H70120 (NID:g1040326) ys01a09.rl"	misc_feature
3599..4017	
/note="similar to EST H89808 (NID:g1080238) yu82g03.rl"	misc_feature
3741..3928	
/note="similar to EST N56845 (NID:g1200735) yw83e04.rl"	misc_feature
3816..3932	
/note="similar to EST AW523936 (NID:g7166321)"	misc_feature
3819..3932	
/note="similar to EST AW503416 (NID:g7139965)"	misc_feature
3820..3932	
/note="similar to EST AW197058 (NID:g6476288) xb15f09.xl"	misc_feature
3820..3932	
/note="similar to EST AW163398 (NID:g6302431) au94g04.yl"	misc_feature
3871..3932	
/note="similar to EST AI700706 (NID:g4988606) we39e12.xl"	misc_feature
3876..3932	
/note="similar to EST T31784 (NID:g613882)"	misc_feature
3889..4387	
/note="similar to EST T79157 (NID:g697666) yd70b07.sl"	misc_feature
4109..4387	
/note="similar to EST AI696334 (NID:g4984234) tw59e05.xl"	repeat_region
4502..4812	
/rpt_family="Alu"	repeat_region
5041..5087	
/rpt_family="Mariner"	misc_feature
5254..5330	
/note="similar to EST AA794004 (NID:g2856959) vr37a02.rl"	misc_feature
5265..5316	
/note="similar to EST AW503416 (NID:g7139965)"	misc_feature
5265..5330	
/note="similar to EST AI700706 (NID:g4988606) we39e12.xl"	misc_feature
5265..5330	
/note="similar to EST AW197058 (NID:g6476288) xb15f09.xl"	misc_feature
5265..5330	
/note="similar to EST AW523936 (NID:g7166321)"	misc_feature
5265..5330	
/note="similar to EST T31784 (NID:g613882)"	misc_feature
5265..5348	
/note="similar to EST AW163398 (NID:g6302431) au94g04.yl"	misc_feature
5284..5330	
/note="similar to EST T30644 (NID:g612742)"	misc_feature
5305..5330	
/note="similar to EST AW993840 (NID:g8254025)"	repeat_region
5750..6041	
/rpt_family="Alu"	misc_feature
5964..6127	
/note="similar to EST AA078486 (NID:g18379660)"	misc_feature
6032..6370	
/note="similar to EST AI566941 (NID:g4525393) tn25b07.xl"	

```
repeat_region 6572..6878
/rpt_family="Alu"
misc_feature 6637..6883
/nc="similar to EST AW748900 (NID:g7663832)"
misc_feature 6884..7119
/nc="similar to EST AW895654 (NID:g8059859)"
misc_feature 6906..7117
/nc="similar to EST AA601344 (NID:g2434969) no16a03.si"
repeat_region 6939..6985
/rpc_family="MIR"
repeat_region 7127..7218
/rpt_family="L1"
repeat_region 7219..7425
/rpt_family="Alu"
repeat_region 7426..7476
/rpc_family="L1"
misc_feature 7510..7697
/nc="similar to EST AA985239 (NID:g3163764) am80d09.si"
misc_feature 7514..7697
/nc="similar to EST T31784 (NID:g613882)"
misc_feature 7514..7697
/nc="similar to EST AW523936 (NID:g7166321)"
misc_feature 7514..7697
/nc="similar to EST AI700706 (NID:g4988606) we39e12.x1"
misc_feature 7514..7689
/nc="similar to EST T30644 (NID:g612742)"
misc_feature 7514..7695
/nc="similar to EST AW197058 (NID:g6476288) xb15f09.x1"
misc_feature 7514..7697
/nc="similar to EST AW993840 (NID:g8254025)"
misc_feature 7514..7697
/nc="similar to EST AA794004 (NID:g2856959) vr37a02.r1"
misc_feature 7565..7697
/nc="similar to EST Z43001 (NID:g570579)"

Query Match 98.4%; Score 3306.6; DB 9; Length 230552;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 3340; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 1 aaaaagtcaggcaactgtggctcattcgttaataccagcaactttggaggccgagggcgt 60
DB 57211 AAAAGGTCAGGCACGTGGCTCATCCGTGTAATCCAGCACATTTGGGAGGCCGAGGCGGT 57270
QY 61 tgaatcacctgaggtcagaggtcagaccagcctgaccaatcgtgaaacctccatc 120
DB 57271 TGAATCACCTGAGGTCAGGAGTTCAGACCAGCCTTGACCAATATCGTGAAACTCCATCTC 57330
QY 121 tactaaaaatacaaaaattagccaggtgtgtgtggcgggtgtctttagtccagcactctg 180
DB 57331 TACTAAAAATACAAAATTAGCCAGGTGTGTGGGGGTGCTTGTAGTCCCAGCTACTTG 57390
QY 181 ggagctgagggcaggtgaattactgaacctgggaggtggaggttgcaatgagccaaagt 240
DB 57391 GGAGGCTGAGGCAGGTGAATTAATTAACCTGGGAGGTGGAGGTTGCAATGAGCCAAAGT 57450
QY 241 tgcacactgactccagtgcagagcagagactccatctcaaaaaaataaaaaaagaag 300
DB 57451 TGCACCACCTGCACTCCAGTGACAGAGCGAGACTCCATCTC-AAAAAATAAAAAAAGAG 57509
QY 301 ttgggaaaaagggcaggtgagtggtccacgacctglaalcccacacactttaagaggtcga 360
DB 57510 TTGGGAAAAGGCCAGGTGCAGTGGCTCCACACCTGTAATCCCAACACTTTAAGAGGCTGA 57569
QY 361 ggtggagaatctttgagccagaggttcagaccagcctggcattgtcccaagaact 420
DB 57570 GGTGGGAGAATCCTTTGAGCCAGGAGTTCGAGACAGCCTGGGCAATTTGCCAAGACCT 57629
QY 421 tctcttaaaaaaattagccgggtgtgtgtgcatcacgtctgtgtccagctattcggg 480
DB 57630 TGCTTTACAAAAAATTAGCCGGGTGTGGTGCATACGTCTATGTGTCCAGCATATTCGGG 57689
QY 481 aggcctgagggcagggagattgcttgagccttaggagctctagtgagctgtgata 540
DB 57770 GAGTCTGGGCTGGTGTGCTGGCTTTGGCAGAAAAAGTCAGGGCTAAGATCTCATCTGACTCTGG 58299
```

```
DB 57690 AGGCTGAGGCAGGGAGATTGCTTGAGCCCTAGAGTCCAGGCGCTGTAGTGTGTGTATCA 57749
QY 541 cgtcacttactcttagcctgggcaacagagcaactctgtctccaaaaaagaataaaa 600
DB 57750 GGTCACTTACTTACTAGCCTGGGCAACAGAGCAAGACTCTGTCTCCAAAAAAGAAATANA 57809
QY 601 gttgggaaggctcactaacttcatcagatgagaacaagacatgtttgaagtgtgaggc 660
DB 57810 GTTGGAAAGGCTCACTAACTTCATCAGATGAGAACAGACATGTTTGAAGTGTGAGGC 57869
QY 661 cgaagcctggagaacgtctatgcgcccaggaatgcagggcagcagagactcaaatgcca 720
DB 57870 CGAAGCCTGGAGAACGCTATGCCCCAGGAAATGCAGGCGACAGAGACTCAAGATGCCA 57929
QY 721 ggcctgttctggaggccagatgggcccctgcaatgcccactcacccljccctccctt 780
DB 57930 GCGCCTGTCTTGGAGGCCAGATGGGCCCTGCAATGCCCACTCACCCCTGCCCTCTCTT 57989
QY 781 gccccagacatcacccggcccctcatcctctgcagacgttaccgcgcatcgccagctacgag 840
DB 57990 GCCCCAGACATCACCGGCCCATCATCTGCAGAGGTACCGGCCCATTTGCCAATACGAG 58049
QY 841 aagacctgggctccagatggctctgtccacgggggacgtgtgtggaggtcgtggagaag 900
DB 58050 AAGACCTCGGCTCCGAGATGGCTCTGTCCACGGGGGACGTGGTGGAGGTCGTAGAGAAG 58109
QY 901 agcaagagcgtcagacctccaccttacggggtccctccctgggtgctcagaaccca 960
DB 58110 AGCGAGAGCGGTACAGACCTCCACCTTACGGGGTCTTCTCCCTGGTGGTCTCAGGAACCA 58169
QY 961 cagccacaagccccctgcgaaggtcagcgagcctggccccctgggaggaactccagctctg 1020
DB 58170 CAGCCACAAGCCCCCTGCCAAGGCTCAGCAGCCTTGCCCCCTGGGAGGACTCCGSCCTCG 58229
QY 1021 ttaggggccctaaatgtctctcccaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
DB 58230 TTAGGGGCCCTAAATGTCTCTCCACACTGTGTGGTGTGGCTTCTGTCTTGTAGTGTGACCC 58289
QY 1081 gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
DB 58290 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 58349
QY 1141 ctggatgggtatgggacgctgttccattatgaagtgggctcagagctgtgattctgtga 1200
DB 58350 CTGGATGGGTATGGGACCGCTGTGTTCATATGAAGTGGGCTCAGAGCTGTGATCTGTGA 58409
QY 1201 gcatgtgtcatgtcatgtgacctcattgtccagtgctgtgtgtgtgtgtgtgtgtgtgtgt 1260
DB 58410 GCATGTGTGCATGCATGCATGTGACCTCACTTGTCCAGGTGTGGTGAAGGTGACATTTCCAA 58469
QY 1261 atctgagcattggacatcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
DB 58470 ATCTGAGCATTTGGACATCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 58529
QY 1321 cagggtgagcgtgtggccctgtccctcagtcacattcccgcacctctgtgtgtgtgtgtgtgt 1380
DB 58530 CAGGGAGCGCTGGGCCCTTGCCCTCAGTCACATTTCCCGCACCTCTGGCACAGGTTGGTG 58589
QY 1381 gttctgtcagatgaaagcaagcagaggtgtggtccagatcctcctcgagccctcgtga 1440
DB 58590 GTTCTGTGATGAAGCAAGCGAGGTGGATCCACGCTCTTCCCTCGAGCCCTCGA 58649
QY 1441 cagtcctgacgagcaggaagacctgagcccaactatgcaggtgtccctgtccctccag 1500
DB 58650 CAGTCTCTACGACGCGAAGACCTTGAGCCCAACTATGCGAGTGGCCCTGCCCTCCGAG 58709
QY 1501 gctgtagggtgtgtggagaagggtgagggcaggggtcaggagatatgtgagtgactgtgtg 1560
DB 58710 GCTGTAGGGGTGTGGAGAAAGAGGGGACGAGGGGTCTAGGGGATATTGAGTGTACTGCTTG 58769
QY 1561 gagtcgtggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1620
DB 58770 GAGTCTGGGCTGGTGTGCTGGCTTTGGCAGAAAAAGTCAGGGCTAAGATCTCATCTGACTCTGG 58829
```

QY 1621 cttggggccctggcagggttgatgcccttggtctggacagggaaacacagggaggagc 1680
|||||
Db 58830 CTTGGGGCCCTGGCAGGTTGTGATGCCCTTGGTCTGGACAGGAAACACAGGAGGAGC 58889
QY 1681 agacgactcggagagtgaggagccagtggtctctgtgatatgtgcccaggttcagtg 1740
|||||
Db 58890 AGACCACCTGGGAGAGTGGGAGGCCAGTGGTCTCTGTGATATGTGGCAGGTTTCAGTGG 58949
QY 1741 gaagctgaaggtatgacagaccttaggtccaggaaggagggctgcctgaaagtgggcca 1800
|||||
Db 58950 GAAGCTGAAGGATGACAGACACCTTAGGCTCAGCAAGAGGGGCTGCTGGAAGTGGGGCA 59009
QY 1801 tcatcactgaccagaaagggaaactggcagtcgagggcagtgccagggctggatgggcccctgcattgag 1860
|||||
Db 59010 TCATCCTGACCAGAAAGGGAACATGGCAGTGCAGGGCTGGATGGGGCCCTGCATTGAG 59069
QY 1861 cttgaaaaaactaataagaattggttaccattttattttattttattttattttattt 1920
|||||
Db 59070 CTTGAAAAAACTAATAGAAATTGGTTACCAATTTTATTTTATTTTATTTTATTTT 59129
QY 1921 tactttttgagatagagctcaactcccttgcctccaggtcaagtgagtgagtcggtatctc 1980
|||||
Db 59130 TACTTTTTGAGATAGAGTCTCAGTCCCTTGTCTAAGGCTGGAGTGGGTGGTGTATCTC 59189
QY 1981 agctcaactgcaacctctgcctccaggtcaagtgatctctccagcctcagcctcccccagg 2040
|||||
Db 59190 AGCTCACTGCAACCTCTGCCTCCAGGATCAAGTCAATCTCCAGCCTCAGCCTCCCAAG 59249
QY 2041 tagctgggattacaagcatgaaccacactgcctgataaatttttttttttttttttt 2100
Db 59250 TAGCTGGGATTAACAAGCATGCAACCACTGCTGCTGATATTTTGTATTTTGTAGTTGAGA 59309
QY 2101 cgggttttcacagcttgccagagctgctcgaacttctcaactcaggtatctgcctg 2160
Db 59310 CGGGTTTCACCAAGTTGGCCAGACTGGTCTCGAACTTCTGACCTCAGGTGATCTGCCGTG 59369
QY 2161 cctcgccctcccaagtgctggaattacagatgtgagccactgtccctggcctggttacc 2220
Db 59370 CCTCGGCCCTCCCAAGTGTGGAATACAGATGTGAGCCACTGTCCTTGGCCTGTTTACC 59429
QY 2221 cacattttaaatgagtgatttacccttttatgtgatttataagcttgtttttttttt 2280
Db 59430 CACATTTTAAATGAGTGATTTTACCCCTTTATGTGATTTACAGCTGTGTTTTTTTTT 59489
QY 2281 ttttttgacaaagtctggctctgtctcccccaggtgagtgagtgagtaaatcagctcagc 2340
Db 59490 TTTTGTGACAAAGCTTGGCTCTGTCAACCCAGGCTGGAGTGCAGTAAATGCAATCTCAGC 59549
QY 2341 tcaactgcaaccttagctcctctgggtttcaagcaattctcctcagccactgagtgagc 2400
Db 59550 TCACCTGCAACCTTAGCCCTCCTGGGTTCAAGCAATTTCTCTGCTCAGCCACCTGAGTAG- 59608
QY 2401 ctggggttacaggaatgacacacacagcagcaggttaatttttttttttttttttt 2460
Db 59609 CTGGGGTTACAGGCATGACCAACACCGCAGGCTAAATTTTGTATTTTGTATTTTGTAGAGATG 59668
QY 2461 ggggttcgccattgtggccaggtgctcgaactcctcagcctcaggtgataccgcccacc 2520
Db 59669 GGGTTTCGCATGTGGCCAGGCTGGTCTCGAACTTCTGACCTCAGGTGATCCGCCGCCG 59728
QY 2521 ttggcctcccaagtgctagattacaggttggaaccacacttgcacagcctgtggtctatc 2580
Db 59729 TTGGCCTCCCAAGTGTAGGATTAAGGTTACAGGTGGGAACCACTCGCCACGCTGTGCTATC 59788
QY 2581 gtttaaacactgggaagcctcagccccccagggccagacagtagctgagctgagcagtt 2640
Db 59789 GTTTAAACACTGGGAAGGCTGACGCCCCCAGCGGACAGTGTAGTGTGAGCTGAGCAGTT 59848
QY 2641 ccagtgccaggtacaggtatgctccaccacacttactatgctgctatcttctgcatagt 2700
Db 59849 CCCAGTGCCAGGTACAGGATGCTCCAGCCACCTACTCATGGCTGATCTCTGTGTCATAGT 59908

QY 2701 gaagtgtcggacagacccttcattcgttatgggatctctggtccccagagtggttgccaat 2760
|||||
Db 59909 GAAGTGTCTGGACACACCTTCATCTTATGACATCTCTGGTCCCCCAGAGTGGGTGGCAAT 59968
QY 2761 gaatgggaatgacaaagctcactcctgggtgtatgggggacagagggccgaagtcacagtgta 2820
|||||
Db 59969 GAATGGGATGGACAAAGCTCACCTGGGTGTAGGGGCGCAGAGGGCCCAAGTCCAGAGTGTA 60028
QY 2821 cccccagagtggttccacagcagagcttgcacagagatctcggatggacagagaggtg 2880
|||||
Db 60029 CCCCCAGAGTGGTCCACAGAGAGCTTGCCAGAGGATCTGGGATGGACAGAGGGTGG 60088
QY 2881 agggaggagaccacagagggggaactgtggccctgggtcgtggtctggagtcctggag 2940
|||||
Db 60089 AGGGAGGAGACCCAGAGAGGGGAACTGTGGCCCTGGGTGGGTCTGGAGTGCCTGGAG 60148
QY 2941 gaagcccaggccagagaggaggaagatgggatgggtggcagccccagcctggggccgacc 3000
|||||
Db 60149 GAAGCCCAAGGCCAGAGAGAGATGGATGGGTGGCGAGCCCGCAGCTGGGCCGACC 60208
QY 3001 tcacactgtctctgtgccccctgcctggaccaggtgagccatcagctcgcacatcagacc 3060
|||||
Db 60209 TCACACTGTGCTCTGTGCCCTGGCCGTGACCCAGGTGAGCCATACGTGCCCATCAAGGCC 60268
QY 3061 tacactgtgtgaggggagagaggtgtccctgtcgcaggggtgaagctgttgagtaatt 3120
|||||
Db 60269 TACACTGTGTGGAGGGGACGAGGTGTCCTGCTCGAGGTTGAAGCTGTTGAGGTCTAT 60328
QY 3121 cacaagctcctggacgggtgtgggtcactcaggttagggagggccccctctccatccagagca 3180
|||||
Db 60329 CACAAGCTCTGGACGGTGGTGGGTCTCATAGGTAGGAGGGCCCTCTCCATCCAGAGCA 60388
QY 3181 ccactctcagtcagcccccagcagcagcaggggttttaggagatctgggtgacttgcct 3240
Db 60389 CCATCTGAGTCAGCCCCCAGCAGGACGGGGTGTTTAGGGATCTGGGGTGACTTGTCCCT 60448
QY 3241 gggactctgggttaagccactgccccctctcctggcttagtttccatctcagtcagcagggag 3300
|||||
Db 60449 GGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGAG 60508
QY 3301 ggatgagccacacttgcctgtcttctgtggggatccaaatgctccttccaaagtcgggtgcat 3360
|||||
Db 60509 GAATGAGGCCACCCCTGCTGTCTTGTGGGATCCAATGTCTTGTCCAAGTGGGTGCAT 60568
QY 3361 t 3361
Db 60569 T 60569

RESULT 3
AC004883
LOCUS AC004883 131359 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP4-771P4 from 7q11.21-q11.23, complete sequence.
ACCESSION AC004883
VERSION AC004883.2 GI:4263746
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131359)
AUTHORS Sulston, J. E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 131359)
AUTHORS Kalicki, J. and Laplant, Y.
TITLE The sequence of Homo sapiens PAC clone RP4-771P4
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 131359)
AUTHORS Waterston, R.H.
TITLE Direct Submission

JOURNAL

Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE

4 (bases 1 to 131359)

AUTHORS

Waterston, R.

TITLE

Direct Submission

JOURNAL

Submitted (20-FEB-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

5 (bases 1 to 131359)

AUTHORS

Waterston, R.

TITLE

Direct Submission

JOURNAL

Submitted (18-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE

6 (bases 1 to 131359)

AUTHORS

Waterston, R.

TITLE

Direct Submission

JOURNAL

Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 24, 1999 this sequence version replaced gi:3309099.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: saplens@watson.wustl.edu

----- Summary Statistics

Center project name: H_DJ0771p04

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send
<mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by
Pieter de Jong and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>) using the method described by
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from
one male donor.

The clone may be obtained either from Genome Systems, Inc.
(<http://www.genomesystems.com>) or Research Genetics, Inc.
(<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: PCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-1186P10. Actual start of
this clone is at base position 1 of RP4-771P4; actual end is at
131359 of RP4-771P4.

FEATURES

source

Location/Qualifiers

1..131359

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q11.21-q11.23"

/clone="RP4-771P4"

/clone_lib="RPCI-4"

repeat_region

1..229

misc_feature

87..549

misc_feature

306..810

misc_feature

373..854

repeat_region

634..945

repeat_region

1111..1407

misc_feature

1649..2063

misc_feature

2043..2408

repeat_region

2276..2945

repeat_region

3052..3361

repeat_region

3424..3947

misc_feature

3445..3819

misc_feature

4350..4703

repeat_region

4716..4992

repeat_region

5005..5265

repeat_region

5268..5501

repeat_region

5522..5828

repeat_region

6359..6671

repeat_region

6697..7001

repeat_region

6740..6981

misc_feature

7127..7471

misc_feature

7567..7859

repeat_region

7884..7937

repeat_region

7938..8101

repeat_region

8381..8683

misc_feature

8418..8752

misc_feature

8602..8708

misc_feature

8603..8708

misc_feature

8604..8708

misc_feature

8604..8708

misc_feature

8607..8708

misc_feature

8610..78320

gene

/gene="GTF2I"

join(8610..8708,10451..10589,18436..18570,19722..19905,

20081..20109,24648..24702,25874..25917,30516..30593,

34336..34395,36373..36429,38357..38419,48276..48386,

49719..49784,51970..52153,53415..53473,54953..55024,

55992..56175,57537..57595,62939..63010,64249..64432,

65374..65432,65828..65902,67505..67606,68515..68580,

68758..68941,70845..70900,71574..71654,72578..72661,

72661..72661)

```
73330. 73513.74993. .75021.76304. .76345.77444. .77485,
78262. .78320)
/gene="GTF21"
/notes="match to AF035737 (PID:g2827180); H_DJ0771P04.1"
/codon_start=1
/product="general transcription factor 21"
/protein_id="AAD15419.1"
/db_xref="gi:4263747"
/translation="MAOVAMSTLPVEDEESSEVMVTFILMSALESCKELAKSAEV
ACIAVETDVFVGTGTRAFVNTKDFDKVYCVEEERKAEAEHMKHSTTQANRM
SVDAVELETUKTVEDYFCYCKALGKSTVVPYKELMDQSAVVVQGLPEGVAK
HPNVDLALIKWILENAGVTSIFIKRFLPEKXHVGRVWMTDADRSLLSPGSGCPI
KVTPEPDSGISLEMAAVTVKEESDPDIYXNIQAGPSDETDDVDEQPLSKLQGS
HSESENEGTEMEYPAEDSTQVPESTSEDEVEVETIEDDDYSPSKRPKANELPQP
VPEPANAGKRKRVENFEKQHNARITDLRKQVEVETFERKYAQAIKAKGVTIPYPLFOS
HEDLYVVEGLPEGIPFRPSPYIPRLERILLAKERIRFVKIKHELLNSTREDILODK
PASVKBEWYARITKLRKMDQDLFCFKAELALGSTEAKAPYQKFEAPHPNDLYVEGLP
ENIFRPSWTFGIPLREKLIQVGNRIKFIKRPPELLTHSTTEVTPRINTPVKEDWN
RITLKRQVEEIFNLKAQALGLTEAVKVPYVVFESNPFLYVEGLPEGIPFRSPYTF
GIPRLERIVRGSNKIKFVYKPELVI SYLPPGMASKINTKALQSPKRPSPGSKVP
FNEKCGALGLKQAVKVPFALFESFPEDFVVEGLPEGVPRPSPSTFGIPREKILRNK
EIVTEGPNNNPOTSAVRPTOTNGSNVFKPRGREFSFEAWNAKITDLKOKVENL
AKIFILIKKPEFETAKESTSSKSPRKINSSPNVNTASGVEDLNLIIQVTFDDDN
ERLSKVRQRLQVNDLFLRKEGEAIGMPVVKYPIKTIINPQCVVVDGMPPGVS
FKAPSYLEISMRRLDLSAEFIKFTVIRPPFLVNNQLVDQSESGEPVIOESAEPQS
LEVPAEIEIKETDSSQIKQEPDPTW"
/misc_feature
/gene="GTF21"
/notes="match to EST AA490549 (NID:g2219722) aa42e07.rl"
8640. .8824
/gene="GTF21"
/notes="similar to EST AA609881 (NID:g2458309) af08d04.s1"

Query Match      98.3%; Score 3304.2; DB 9; Length 131359;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 3339; Conservative 0; Mismatches 18; Indels 4; Gaps 2;

Qy      1  aaaaagtcaggcactgtgctcatcctgtaatccacagcacttggggagccgagcggt 60
Db 101664  AAAAGGTGAGGCACCTGTGGCTCATCGCTGTATCCAGCACTTTGGGAGCGCGAGCGGT 101723

Qy      51  tgaatcactgaggtcaggagtcagaccagcctgaccaaatactgtaaacctccatctc 120
Db 101724  TGAATCAGCTGAGGTTCAGAGTTCGAGACGACGCTGACCAATATCGTAAATCCATCTC 101783

Qy      121  tactaaaaatacaaaattagcaggtgtgtgagcggtgctgtagtgtccagctactctg 180
Db 101784  TACTAAAAATACAAAATTAGCAGGTGTGTGGCGGGTGTGTGTAGTCCCAAGCTACTTG 101843

Qy      181  ggaggtgagcgagtgtaattacttgaacctggggaggtggaagttgcaatgagccaaagt 240
Db 101844  GGAGGCTGAGCGAGGTGAATTACTTGAACCTGGGAGGTGGAGGTGCAATGAGCCAAAGAT 101903

Qy      241  tgcaccactgcaactcagtgacagagcagactccatctcaaaaaaataaaaaaaag 300
Db 101904  TGCACCACTGCACTCCAGTGACAGAGCAGACTCCATCTC---AAAAAATAAAAAAAG 101960

Qy      301  ttgggaaaagccaggtcagtggtccacgcctgtaatcccaaacactttaagaggtcga 360
Db 101961  TTGGGAAAAGGCCAGGTGCAGTGGCTCCACGCTGTAAATCCCAACACTTTAAGAGGCTGA 102020

Qy      361  ggtggagaactcttgagccagaggtctcagagccagcctgggagcttgcctccagagcct 420
Db 102021  GGTGGAGAAATCTTTGAGCCAGGAGGTTCGAGACCAAGCTTGGGGCATTTGCCAAGACT 102080

Qy      421  tgtctttcaaaaaataatagccgggtgtggtggcatactctgtgtgtccagctattcggg 480
Db 102081  TGTCTTTACAAAAAATTAGCCGGGTGTGGTGCATACGCTATATGTGTCCAGCTATTCGGG 102140

Qy      481  aggcctgagggcagggagatgtcttgagccttagagtcctagggcctgtagt-gagctgtga 540
Db 102141  AGGCTGAGGCAGGGAGATTTGCTTGAAGCTAGGAGTCCAGGGGCTGTAGTGAGCTGTGATCA 102200
```

```
Qy      541  cgtcactgtactctagcctggggaacagagcaagactctgtgtcccaaaaaaataaaa 600
Db 102201  CGTCACTGTACTCTAGCCTGGCAACAGAGCAAGACTCTGTCTCCAAAAAAGANAATAA 102260

Qy      601  gttggaaaagctcactaacttcaatagatgagaacaagacatgtttgaaagtgtgaggc 660
Db 102261  GTTGGAAAGGCTCACTAACTTTCATCAGATGAGAACAAAGGACATGTTTGAAGTGTGAGGC 102320

Qy      661  cgaagccttggaagacgtctatgcgcagagaaatcagggcagcagacagatgaatgcca 720
Db 102321  CGAACCTTGGAAGACGCTATGCGCCAGAAATGAGGGCAGAGACAGACTCAAGATGCCA 102380

Qy      721  ggcctgttctggagccagatgggcccctgcaatgcccactcaacctgacctctctct 780
Db 102381  GGCCTGTCTTGAGGCCAGATGGCCCTGCAATGCCACTCACCTTGCCCTCCCTCTCTT 102440

Qy      781  gccccagacatacaacggcccccatactctctctgagacgttaccgcccattgcccactacgag 840
Db 102441  GCCCCAGACATCACCGGCCCATCATCTCTGACAGAGTACCGGCCATTGCCAATACGAG 102500

Qy      841  aagacctcgggctccgaatagctctgtcccacggggagcgtgtgaggtcgtgagaag 900
Db 102501  AAGACTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCTGCTGAGGCTCTGATAGAG 102560

Qy      901  agcgagagcggtcagacacctcccacacctacggggctcctctccctggtgctcaggaacca 960
Db 102561  AGCGAGAGCGGTACAGACTCCACCTTACGGGGCTCTTCCCTGGTGTCTCAGGAACCCA 102620

Qy      961  cagccaaagcccccctgcgaaggtcagggcctcagggcctggcccctgggagactccagctctg 1020
Db 102621  CAGCCACAAAGCCCCCTGCCAAGGCTCAGCGACGCTTGGCCCTGGGAGGACTCCGCTCTG 102680

Qy      1021  ttaggggccctaaatgtctcccccacactgtgggtcgctctctctcttagtgtcacct 1080
Db 102681  TTAGGGGCCCTAAATGTCTCCCCACACTGTGGGTGCGCTCTGTCTTAGTGTGACCCCT 102740

Qy      1081  gtggtggctgtgggcatctgtgcatggcagggccggggcgggcgtctctgctgtctctgt 1140
Db 102741  GTGTTGGCTGTGGGCATCTGTGCATGGCAGGCGGGGCGGGGCATGTCTGCTGTCTGT 102800

Qy      1141  ctggatgggtatgggaccgctctgttcataatgaagtgggctcagagcgtgtgattctgtga 1200
Db 102801  CTGGATGGGTATGGGACGCTCTGTTTCATTATGAAGTGGGCTCAGAGCTGTGATCTGTGA 102860

Qy      1201  gcattgtgcacatgcacatgtgacctcattgtccaggtgtggtgaggtgacatttccaa 1260
Db 102861  GCATTGTGCATGCATGCATGTGACCTCATTTGTCAGTGTGGTGAAGGTGACATTTCCAA 102920

Qy      1261  atctgagcattggacatcagtgatgtctgtgtccctgtgtccctcacatccctcgtgagcgtg 1320
Db 102921  ATCTGAGCATTTGGACATCAGTGTGTGTGTGTCCTGTCTCCTCACCATCCCTGATGGCTG 102980

Qy      1321  cagggagccgctgggcccctgcccctcagtcacattcccgacacctctggcacaggttggtg 1380
Db 102981  CAGGGAGCCGCTGGGCCCTGCCCTCAGTCACTATTCGCGACCTCTGGCACAGGTGGTG 103040

Qy      1381  gttctgtcagatgaaagcaaaagcagaggtggtatcccgacatcccttccctcagagccctcga 1440
Db 103041  GTTCTGTCAAGTGAAGCAAAAGCGAGGTGGATCCACGCGTCTTCTCCGAGCCCTCGA 103100

Qy      1441  cagtcctgacagagcagagaccctcgagcccaactatgcaggtgccccctgccccccag 1500
Db 103101  CAGTCCTGACGAGACGGAAGACCTTGAGCCCAACTATGCGAGGTGCCCTGCCCTCCGAG 103160

Qy      1501  gctgtagggggtgtgggaaaggggcaggcaggggtcaggggatattgagtgactgctttg 1560
Db 103161  GCTGTAGGGGTGTGGAGAAAGGGGACGACGGGCTCAGGGATATTGAGTGACTGCTTTG 103220

Qy      1561  gagtcgtggcgtggtgtgcttggcttggcagaaaaagtcagggcctaagatctcatcgtcgtg 1620
Db 103221  GAGTCGGGCTGGTGTGCTGGCTTGGCAGAAAAGTCAGGGCTAAGATCTCATCGACTCTGG 103280

Qy      1621  cttggggggccctggcaggtgtgtgtgtagtgccttggctgtgagacagggaaccagaggagc 1680
```

|||||
Db 103281 CTTGGGGCCCTGGCAGGTTGTGATGCCCTTGCTGTGCACAGGAAACCAGGAGGAGGC 103340
QY 1681 agacagactcggagagtgaggagccagtggtctctgttgatattgtgcccaggttcagtg 1740
Db 103341 AGACAGCTGGGAGAGTGGGAGCCAGTGGTGTCTGTGTGATATGTGGCAGGTTTCAGTGG 103400
QY 1741 gaagctgaagagtagagcagacaccttaggtccaggaagaggggcctgcctggaagtggggca 1800
Db 103401 GAAGCTGAGGATGAGCAGACCTTAGGCTCAGGAAGAGGGGCTGCTGGAAGTGGGGCA 103460
QY 1801 tcatcactgaccagaaagggaaacctggcagtgccagggctggatggggccctgcattgag 1860
Db 103461 TCATCACTGACCAGAAAGGAAACTGCGAGTGCAGGGCTGGATGGGCGCTGCATTGAG 103520
QY 1861 cttgaaaaaactaataagaattggttaacattttttattattattattattattattatt 1920
Db 103521 CTTGAAAAAATAAATAGAAATGGTTACCAATTTATTTATTTATTTATTTATTTATTT 103580
QY 1921 tactttttgagatagagtctcaactcccttgctaaggctggagtgcggtggtgctatctc 1980
Db 103581 TACTTTTTGAGATAGATCTCACTCCCTTGCTAAGGTTGGAGTGGGTGCTATCTC 103640
QY 1981 agtcaactgcaacctctgctctccaggatcaagtgtattctccagcctcagcctcccccagg 2040
Db 103641 AGTCACTGCAACCTCTGCCTCCAGGATCAAGTGATTCTCCAGCCTCAGCCTCCCAAG 103700
QY 2041 tagctgggattacaagcatgacacacatgcctggatgcgtgggtgctatctc 2100
Db 103701 TAGCTGGGATTACAAGCATGACACCACTGCGCTGGATTAATTTTATTTTATTTTATTT 103760
QY 2101 cgggggttcacaggttgggcagactggtctcgaactctcgaacctcagtgatctgcctg 2160
Db 103761 CGGGGTTTCACAGGTTGGCCAGACTGGTCTCGAATCTTGACCTCAGGTGATCTGCCGTG 103820
QY 2161 cctcggcctccaaagtctgggaaattacagatgtagcgaactgtccctggcctggttacc 2220
Db 103821 CCTCGGCCTCCCAAAAGTGTGGAATACAGATGTGAGCACTGTCCCTGGCGTGGTTACC 103880
QY 2221 cacatttaaaatgagtgattctaccccttttatgtgatttaacagctgttttttttt 2280
Db 103881 CACATTTAAATGAGGTGATTTTACACCTTTTATGTGATTTTACAGCTGTGTTTTTTT 103940
QY 2281 ttttttgacaaaagtctggctctctcaccaggctggagtgagtaaatgcaatctcagc 2340
Db 103941 TTTTTTGAGACAAAGCTTGGCTCTGTCACCCAGGCTGGAGTGCAGTAATGCAATCTCAGC 104000
QY 2341 tcaactgcaaccttagcctcctgggttcaagcaattctcctcagccacactgagtagc 2400
Db 104001 TCACCTGCAACCTTAGCCCTCTGGGTCAAGCAATTTCTCTGCTCAGGCACCTGAGTAG - 104059
QY 2401 ctgggggttacagcgtgacaccacgcagcgtcaattttttgtatttttagtagagatg 2460
Db 104060 CTGGGGTTACAGCATGCACACACGCGCAGGCTAATTTTTTGTATTTTATGATGATG 104119
QY 2461 ggggttcgcattgtggccaggctgggtctgaactccctgacctcagctgagtgatccggccgc 2520
Db 104120 GGGTTTGGCCATGTTGGCCAGGCTGGTCTCGAACTCTTGACCTCAGGTGATCCGCCCGCG 104179
QY 2521 ttggcctcccaagtcttaggattacaggtgggaacacaccttgcccagccttggtctatc 2580
Db 104180 TTGGGCTCCCAAAAGTGTAGGATTAAGGTGGGAACCACTCGCCAGCCTGTGGCTATC 104239
QY 2581 gtttaaacactgggagggcctgacgccccccagggccgacagttagctgagctgagcagtt 2640
Db 104240 GTTTAAACACTGGGAGGGCTGCAGCCCCAGCCGACAGTTAGTGTGACGTGACAGTT 104299
QY 2641 ccagtgccaggtagacagtagctccacccactactcatggtgtagctctttgtgctatg 2700
Db 104300 CCCAGTGGCAGGTAGACGGATGCTCCAGCCACCTACTCATGCTGATCTCTGTGTCATAGT 104359
QY 2701 gaagtgtctgacacaccttcatcttatggtatctctggtcccccagagtggtggcgaat 2760
|||||

Db 104360 GAAAGTGTCTGGACAGACCTTCATCGTTATTAGGATCTCTGGTCCCCAGAGTGGTGGCAAT 104419
QY 2761 gaattggagttgagaaagctcaccctgggtgttagggggcagagggccgaagtccagagtta 2820
Db 104420 GAATGGGAGTGGACAAGCTCACCTGGGTGTAGGGGGCAGAGGGCCGAAGTCCAGAGTGA 104479
QY 2821 ccccccagagtggtgtccagcagagagcttgccagagagatctctgggatggagcagaggggtg 2880
Db 104480 CCCCCAGAGTGGTGTCCAGCAGGAGCTTCCCAGAGGATCTGGGATGGAGCAGGAGGGTGG 104539
QY 2881 agggaggagacccagaagaggggaaactgtgggcccctgggtgggtctgtggatgacctggag 2940
Db 104540 AGGGAGGAGAGCCACAAGAGGGGAACTGTGGCCCTGGTGGTGGTGGTGGTGGTGGAG 104599
QY 2941 gaagcccaaggcagagagagaagatgggatgggtggtgcgagagcccccagagctggggcagc 3000
Db 104600 GAAGCCCCAGGCGCAGAGAGGAGAAATGGGATGGGTGGGAGCCCCCAGGCTGGGCCGAGC 104659
QY 3001 tcacactgtctctgtgcccctgcccgtggaccaggtgagccatcagtcgcatcaagggc 3060
Db 104660 TCACACTGTGCTGTGTCGCCCTGCGGTGACAGGTGAGCCATAGTGCGCCATCAAGGCC 104719
QY 3061 tacactgtctgtgggggggacgaggtgtccctgtctcaggtgaggtgtgaggtgtgagtaatt 3120
Db 104720 TACACTGTGTGGAGGGGAGGAGGTGTCCCTGTCTCGAGGGTGAAGCTGTTCAGGTCAAT 104779
QY 3121 cacaagctccctggacgctggtgggtcatcaggttaggagggcccccctccatccagagca 3180
Db 104780 CACAAGCTCTCTGGAGCGGTGGTGGGTCTACAGGTAGGAGGGCCCCCTCTCCATCCAGAGCA 104839
QY 3181 cccatctgagtcagccccagcagcagcgggtgtttagggatctctgggtgagctgtccct 3240
Db 104840 CCCATCTCAGTCAGCCCCAGCCAGCAGCGGGGTGTTTAGGATCTGGGCTGACTTGTCCCT 104899
QY 3241 ggggaactcgtgggtaagcactgcccctctctgtgggttagttccatctcagtagcagggag 3300
Db 104900 GGGACTCTGGGTAAAGCCACTGCCCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAG 104959
QY 3301 ggatagccccacctctgctctgtgggaatccaatgctcttccaaagtgggtgcat 3360
Db 104960 GAATGAGCCACCCCTTGCTGTGTGGGATCCAAATGCTCTGTGCCAAGTGGGTGAT 105019
QY 3361 t 3361
Db 105020 T 105020

RESULT 4

AC083884
LOCUS AC083884 169604 bp DNA PRI 23-MAY-2001
DEFINITION Homo sapiens clone RP11-813J7, complete sequence.
ACCESSION AC083884
VERSION AC083884.6 GI:14190780
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 169604)
TITLE Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 169604)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

Db 106691 TACTTTTTTGTAGATAGAGTCTCACTCCCTTGCTAAGGTTGGAGTGGGTGCTATCTC 106750
QY 1981 acctcaactgcaactctctccctccagagatcaagtatctccagcctcagcctcccccagg 2040
Db 106751 AGCTCAGTCAACCTCTGCCTCCCAAGATCAAGTGATTTCTCCAGCCTCAGCCTCCCAAG 106810
QY 2041 tagctgggattacaagcatgcaccaccatgcctggatatttttttagttgaga 2100
Db 106811 TAGCTGGGATTACAGCATGCACCACTGCTGGATTAATTTTGTATTTTAGTTGAGA 106870
QY 2101 cggggttccacagattggccagactggctctcgaaactctgacctcagtgatcgcctg 2160
Db 106871 CGGGGTTTCCACAGGTTGCCAGACTGGTCTCGAACTTCTGACCTCAGTGATCTGCCCTG 106930
QY 2161 cctcgagctcccaagactgagaaattacagatgtgagccactgtccctggcctgattacc 2220
Db 106931 CTTGGCCTTCCCAAGTGTGGAATTTACAGATGTGAGCACCTGTCCCTGGCGTGGTTACC 106990
QY 2221 cacattttaaaatggagtattccacctttttatggtatttcacagctgttttttttttt 2280
Db 106991 CACATTTTAAATGAGTGATTTTCAACCTTTTATCTGGATTTACAGCTTGTTTTTTTTT 107050
QY 2281 ttttttgagaaaaagtctggtctctgcacccaggtctgagtgagtaataatcagc 2340
Db 107051 TTTTGTGACAAAAGTCTGGCTCTGTACCCAGGCTGGAGTGCAGTAATGCAATCTCAGC 107110
QY 2341 tcactgcaaccttagcctctggttcaagcaattctcctcctcagccacctgagtagc 2400
Db 107111 TCAGTGCAACCTTAGCCTCTGGGTTCAGCAATTTCTCTGGCTCAGCCACCTGAGTAG- 107169
QY 2401 ctggggttacagcatgcaccaccacgcagcgttaatttttggatttttagtagagatg 2460
Db 107170 CTGGGTTTACAGCATGCACCAACCAAGGCTGAGTGCAGTAATGCAATCTCAGC 107229
QY 2461 ggggttcgcatgttggccaggtggtctgcaactcctcagcctcaggtgagtcgcccgc 2520
Db 107230 GGGTTTCGCATGTGTGGCCAGAGCTGTCTCGAATCTCTGACCTCAGGTGATCGCCCGC 107289
QY 2521 ttggcctcccaagtgctaggtatcacgtgggaaccacttgccagcctgtgctatc 2580
Db 107290 TTGGCCTCCCAAGTGTAGGATTTACAGGTGGGAACCACTCTGCCAGCCTGTGCTATC 107349
QY 2581 gtttaaacactgggaagcctcagccccagccagccagcttagctgagctgagcagtt 2640
Db 107350 GTTTAAACACTGGGAAGGCTCAGCCGCCAGCCGACAGTTAGTGCAGCTGACAGTT 107409
QY 2641 ccagtgccagtagagcatgctcaccaccactactcagtgatctcttctgctatagt 2700
Db 107410 CCCAGTGCAGTAGACGGATGCTCCAGCCACTACTATGCTGCTGCTCATAGT 107469
QY 2701 gaagtgtctggacagaccttcatgttatgggatctctggtcccccagagtggtggcaat 2760
Db 107470 GAAGTGTCTGGACAGACCTTCACTGATTTAGGATCTCTGGTCCCCAGAGTGGTGGCAAT 107529
QY 2761 gaatggagtgacaaagctcacctgggtgtaggggagagagccgaagtcacagagtga 2820
Db 107530 GAATGGAGTGCAGAGCTCACCCTGGTGTAGGGGACAGAGGCGCAAGTCCAGAGTGA 107589
QY 2821 cccccagagtggtggtccagcagagagctgcccagggatcttggatggagcagagggctg 2880
Db 107590 CCCCCAGAGTGGTGTCCAGCAGAGCTTCCAGGGATCTGGGATGGAGCAGAGGGTGG 107649
QY 2881 agggagagaccagaaagaggggaactgtggccctgggtgggtctggagtgccctggag 2940
Db 107650 AGGGAGGAGACCCAAAGAGGGGAACCTGTGGGCCCTGGGTGGGTGGAGTGCCTGGAG 107709
QY 2941 gaagccagagccagagagagaaagatggatgggtggcagagcccgagctgggcccagcc 3000
Db 107710 GAAGGCCAGCCGACAGAGAGAGATGGATGGGTGGCGAGCCCCAGGCTGGGCCGACC 107769
QY 3001 tcacactgtgctctgtgcccctgcccgtggaccaggtgagccatacgtgccatcaagccc 3060
Db 107770 TCACACTGTGCTCTGTGCCCCCTGCCGTGGACCAAGGTGAGCCATAGTGCCTCAAGGCC 107829

QY 3061 tacactgtctggaggggacgaggtgtccctgctcagaggtgaagctgttgaggttaatt 3120
Db 107830 TACTACTGTCTGGAGGGGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTGGAGTCAAT 107889
QY 3121 cacaagctcctggagcggctggtgggtcatcaggtaggaggggccctctccatccagagca 3180
Db 107890 CACAAGCTCCTGGACGGCTGTGGTGCATCAGGTAGGAGGGCCCTCTCCATCCAGAGCA 107949
QY 3181 cccatctagtcagccccagccagagcggggtgttttaagggatctggggtgacttgccct 3240
Db 107950 CCCATCTCAGTCAGCCCCAGCCAGACGCGGTGTGTAGGATCTGGGTGACTTGTCCCT 108009
QY 3241 gggactctgggttaagccactgccccctctctgggttagtttccatctcagtagcaggag 3300
Db 108010 GGGACTCTGGGTAAAGCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGAG 108069
QY 3301 ggatgagccaccttgcctgtcttgggtggatccaatgtccttgtccaaagtgggtgcatt 3360
Db 108070 GAATGAGCCCAACCTTGCTGTGTGGGATCCCAATGTCTTGTCCCAAGTGGTGCAT 108129
QY 3361 t 3361
Db 108130 t 108130
RESULT 5
AC004166 AC004166 275197 bp DNA PRI 20-JUL-2000
LOCUS Homo sapiens Chromosome 7 BAC Clone 239c10, complete sequence.
DEFINITION AC004166
ACCESSION AC004166
VERSION AC004166.12 GI:8887011.
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 275197)
AUTHORS Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Homo sapiens Chromosome 7 BAC Clone 239c10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 275197)
AUTHORS Burian,D.M. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 275197)
AUTHORS Burian,D.M., Ren,Q., Meadows,S., Huang,E., Korenberg,J. and
Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 275197)
AUTHORS Ren,Q., Burian,D., Huang,E., Meadows,S., Korenberg,J. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jul 1, 2000 this sequence version replaced gi:8779478.
FEATURES Location/Qualifiers
source 1.275197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="239c10"
BASE COUNT 70874 a 66648 c 66737 g 70938 t
ORIGIN

Query Match 98.0%; Score 3293.2; DB 9; Length 275197;
Best Local Similarity 99.3%; Pred. No. 0;

	Matches	3339:	Conservative	0:	Mismatches	18:	Indels	5:	Gaps	3:
Qy	1	aaaaggtcaggcactgtgctcatgctctgtaatccacgacactttgggaggccgagcggt	60							
Db	43597	AAAAGGTcAGGCACtGTGgCTCATGcCTGTATCCcAGCACtTTGGGAGGCGGAGCGCGT	43656							
Qy	61	tgaatcaactgaggtcagaggttcogagaccagcctgaccaatatatcgtagaaactccatctc	120							
Db	43657	TGAATCACTGAGGTcAGGAGTTCGAGACcAGCCTGACCAATATCGTGAAACTCCATCTC	43716							
Qy	121	tactaaaaatacaaaaaattagccagggtgtggtggcggtgtctgttagtccagactacttg	180							
Db	43717	TACTAAAAAATACAAAAATTAGCCAGGTGTGTGGCGGTGTGTGTAGTCCcAGTACTTTG	43776							
Qy	181	ggaggctgagcgaggtgaattacttgaacctgggaggttgagggttcgaatgagccaagat	240							
Db	43777	GGAGCTGAGcAGGTGAATTACTTTGAACCTTGGAGGTTGGAGGTTCcAATGAGCCAAcAT	43836							
Qy	241	tgcaccactgcactccagtgacagagcgagactccatctcaaaaaaataaaaaaag	300							
Db	43837	TGCACCACTGCACtCCAGTGACAGAGcGAGACTTCCATCTC--AAAAAATAAAAAAAG	43893							
Qy	301	ttggaaaaagccagggtgcagtgcctcacgcctgttaatcccaacacttttaagaggtcga	360							
Db	43894	TTGGAAAAGCCcAGGTGCAGTGCCTCCACGCCCTGTAAATCCCAACACTTTTAAGAGGCTGA	43953							
Qy	361	ggtgggaaatcccttgagcccgagagttcgagaccagcctggagcttgtcccagacct	420							
Db	43954	GGTGGGAATCCCTTTGAGCCcAGGAGTTCGAGACCAGCCTTGGGcATTGTCCcAAGACCT	44013							
Qy	421	tgtctttcaaaaaattagccgggtgtggtggccatacgtctgtgtccagctatttcggg	480							
Db	44014	TGCTCTTACAAAAATTAGCCGGGTGTGGTGGCATACGTCTATGgTCCcAGCTATTTCGG	44073							
Qy	481	aggctgagcgaggagatttgtctgaacctaggagctctagggtctagtgagctgtgacaa	540							
Db	44074	AGGCTGAGGCAGGAGAGATTGTCTTGAGCCTAGAGGTTCcAGGGCTGTAGTGTGATGA	44133							
Qy	541	cgtcacgtactctagcctggcgcaacagagcaagactctgtctcaaaaaagaaaaataa	600							
Db	44134	CGTCACGTACTCTAGCCTGGGCAACAGAGcAGACTCTGTCTCCAAAAAGAAAAATAA	44193							
Qy	601	gttggaaaaagctcactaaacttcacagatgagaaacaaagacatgttctgaagttgtaggc	660							
Db	44194	GTTGGAAAAGcCTCACTAACCTTCATCAGATCAGAACcAGGACATGTTTGAAGTGTGAGGC	44253							
Qy	661	cgaagcttgagaaacgtatgcgccaggaaaaatgcagggcgagagagactcaagatgcca	720							
Db	44254	CGAAGCCTGGAGAACGTATTGCGCCcAGGAANTGCAGGGCAGcAGAGACTCAAGATGCcA	44313							
Qy	721	gcgcctgtcttgaggcccgatggccctgaatgccactcaccctgcctccctctt	780							
Db	44314	CGCGCTGTCTTGGAGGGCCcAGATGGCCCTGCcAATGCCcACTCACCCTTGCCTCTCTT	44373							
Qy	781	gcccagacatacccgcccatactctgcagacgtaccgcgccaattgcccagactacag	840							
Db	44374	GCCCcAGACATcACCGGCCcCATCATCTGcAGAGTACCGCGCCcATTGcCAACTACGAG	44433							
Qy	841	aagacctgggctccgagatggctctgttccacgggggaacgtgggtggaggttcgtggagaag	900							
Db	44434	AAGACCTcGGGCTCCGAGATGGCTCTGTCTCCACGGGGAGcGTGGTGGAGTGTGAGAAAG	44493							
Qy	901	agcgagcggtcagactccacacttacggggtccttcccctggctcgaagaacca	960							
Db	44494	AGCGAGAGCGGTcAGACCTCCcACCTTACGGGGGTCTTCCcCTTGGTGTCTCAGAAACCCcA	44553							
Qy	961	cagccaaagcccctgcgaaggttcaggagcctggcccctgggaggaactccagctctg	1020							
Db	44554	CAGCCAAAGCCcCTGCCAAGGCTCAGGCAGCCTTGCCCTTGGGAGcACTTCCCGCTCTG	44613							
Qy	1021	ttaggggccctaaatctctccccacactgtgggtgcgcctctctctttagtgtcacct	1080							
Db	44614	TTAGGGGCCCTTAAATGTCTCTCCcACACTTGGGTGGCTTCTGTCTTGTAGTGTGCACCT	44673							

QY 1 aaaaagctcaggcaactgtgtgctcatctgctgtataatccacagcaactttggaggccgagcggt 60
|||||
Db 10114 AAAAGGTGAGGCACCTGTGGCTCATGCTGTAATCCAGCACCTTTGGAGGCCGAGCGGT 10173
|||||
QY 61 tgaatcacctgaggtcaggagttcagagaccctgaccaaataatcgtgaaacctccatctc 120
|||||
Db 10174 TGAATCACCTGAGGTTCAGAGTTCGAGACCAAGCCCTGACCAATATCGTGAATCCCATCTC 10233
|||||
QY 121 tactaaaaatacaaaaaattagccaggtgtggtggcggtgtgctgtagtcceagctacttg 180
|||||
Db 10234 TACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCGGTCTGTGTAGTCCCAAGCTACTTG 10293
|||||
QY 181 gaggctcaggcagtgtaattacttgaaacctgggaggttgaggttgcaatgagccaagat 240
|||||
Db 10294 GGAGGCTGAGGCAGGTGAATTAATTGAACCTGGGAGGTGGAGGTTGCAATGAGCCAAAGAT 10353
|||||
QY 241 tcacacactgcactccagtgcagagcagagactccatctcaaaaaaataaaaaaag 300
|||||
Db 10354 TGCACCACTGCACTCCAGTGACAGAGCGAGACTCCCATCTC - AAAAAAANAANAANAAG 10412
|||||
QY 301 ttgggaaaagccaggtgcagttggtccacgctgttaatcccaacactttaagaggtga 360
|||||
Db 10413 TTGGGAAAAGGCCAGGTGCAGTGGCTCCACGCTGTATATCCCAACACTTTAAGAGGCTGA 10472
|||||
QY 361 ggtggagaatcctttgagccagaggttcgagaccagcctgggcaattgtcccaagaact 420
|||||
Db 10473 GGTGGGAGAACTCTTTAGCCCCAGGAGTTGAGAGCAGCAGCTGGGCATTTGTCCAAGACCT 10532
|||||
QY 421 tctcttcaaaaaattagccgggtgtggtggcatacagctctgtgtccagctattcggg 480
|||||
Db 10533 TGTCTTTTACCAGAAATTAGCCGGGTGTGGTGGCATACGTCATATGTCGCCAGCTATTCGGG 10592
|||||
QY 481 aggctgagcagggagattgcttgagccttaggtctcagggctgtagtgagctgtgatca 540
|||||
Db 10593 AGGCTGAGCAGGGAGATTGCTTTGAGCTTAGGCTCCAGGCTGTAGTGAGCTGTGATCA 10652
|||||
QY 541 cgtcactgtactctagcctgggcaacagagcaagactctgtcccaaaaaaataaaa 600
|||||
Db 10653 CGTCACTGTACTCTAGCTTGGGCAACAGAGCAAGACTCTGTCTCCCAAAAAGAAATAAA 10712
|||||
QY 601 gttggaaggtcactaacttcacagatgagaaagacacatgtttgaagtgtgagcc 660
|||||
Db 10713 GTTGGAAAGGCTACTAATCTCATGATGAGAACAGGACATGTTTGAAGTGTGAGGC 10772
|||||
QY 661 cgaagcctggagaaacgctatgcgccaggaataatgcaggcagcagagactcaagatgcca 720
|||||
Db 10773 CGAAGCCTGGACACGCTATGCGCCCGAGGAATGCAGGCGCAGAGACTCAAGATGCCA 10832
|||||
QY 721 gcgctgttctggagccagatgggcccctgcaatgcccactcaacctgcccctctctt 780
|||||
Db 10833 GCGCCTGTTCTGGAGGCCAGATGGGCCCTGCAATGCCACCTCACCCCTGCCCTCCTCTT 10892
|||||
QY 781 gcccagacatcacggcccatactctgcagacgttacccgcccattgcgcactacag 840
|||||
Db 10893 GCCCAGACATCACCGCCCCATCATCTCTGCAGAGCTACCGGCCATTGCCAACTACGAG 10952
|||||
QY 841 aagacctgggctccagatggctctgtccacgggggacgtgtggtgaggtcgtggagag 900
|||||
Db 10953 AAGACCTCGGGCTCCAGATGCTCTGTCCACGGGGAGCTGTGTGAGGTCTAGAGAAG 11012
|||||
QY 901 agcgagagcggtcagacctcccaacttaaggggctctctccctgggtgctcaggaaccca 960
|||||
Db 11013 AGCGAGAGCGGTTCAGACCTTCCCACTTACGGGGCTCTTTCCCTGGTGTCAAGAACCCA 11072
|||||
QY 961 cagccacaagccccctgcaaggtcaggcagcctggcccctggaggagactccagctctg 1020
|||||
Db 11073 CAGCCACAAGCCCCCTGCAAGCTCAGGACGCCCTGGCCCCCTGGGAGGACTCCGGCTCTG 11132
|||||
QY 1021 ttaggggccctaaatgtctctcccaactgtgggtgcgctctctctcttagtgtgcacct 1080
|||||
Db 11133 TTAGGGGCCCTAAATGTCTCTCCCACTAGTGGGTGGCTTCTCTCTTAGTGTGACCCCT 11192
|||||
QY 1081 gtggtgtgtgggcatctgtgcatggcagggccggggggcagtcgtctgtctgttctgt 1140
|||||

Db 11193 GTGTGGCTGTGGGATCTGTGCATGGCAGCGCGGGCGGCATGCTGTGCGTGTCTGT 11252
|||||
QY 1141 ctggatgggtatggacacgtctgttcattatgaagtgggctcagagctgtgattctgtga 1200
|||||
Db 11253 CTGGATGGGTATGGGACCGCTGTTCATTATGAAGTGGGCTCAGAGCTGTGATCTGTGA 11312
|||||
QY 1201 gcatgtgtcatgcatgcatgtagacctcatgtccagttgtgtgaaggtgacatttccaa 1260
|||||
Db 11313 GCATGTGTGCATGCATGTGACCTCATTTGCCAGTGTGTGAAGGTGACATTTCCAA 11372
|||||
QY 1261 atctgagcatctggacatcagttgtgtgtccctgtgtctctcaacctccctgtaggtctg 1320
|||||
Db 11373 ATCTGAGCAATTGGACATCAGTGTGTGTCTCTGTCTGTCTCCTCACCATCCCTGATGGCTG 11432
|||||
QY 1321 caggagcgcctggccctgccccctcagtcacatctccgcacactctggcacaggttggctg 1380
|||||
Db 11433 CAGGAGCCCTGGGCCCTGCCCTCAGTCACATTTCCCGACCTCTGGCACAGTTGGTG 11492
|||||
QY 1381 gttctgtcagatgaagcaagcgaggtctggatccacagcatcttctcagagccccctgga 1440
|||||
Db 11493 GTTCTGTACAGTGAAGCAAAAGCAGGCTGGATCCACAGCATCTTCTCGAGCCCTGGA 11552
|||||
QY 1441 cagtcctgcagagacggaagaccctgagccccactatgcaggtgccccctgccccctcag 1500
|||||
Db 11553 CAGTCTCGACGAGACGGAAGACCTTGAGCCCAACTATGCGAGGTGCCCTCGCCCTCCGAG 11612
|||||
QY 1501 gctgtagggtgtggagaaaggagcagcagggctcagggtatttgactgactcttg 1560
|||||
Db 11613 GCTGTAGGGGTGTGGAGAAAGGGCAGGCGTCAGGGATATTGAGTGACTGCTTTG 11672
|||||
QY 1561 gagtcgggctgtgtgtgctgtggcagaaagtcagggtcagggtcagatctcctcgtctgtg 1620
|||||
Db 11673 GAGTCTGGGCTGTGTGCTGACTTGGCAGAAAAGTCAGGGCTAAGATCTCATCGACTCTGG 11732
|||||
QY 1621 cttgggggccccctggcaggtgtgtgtgtcccttggctggacagggaaacccggggaggagc 1680
|||||
Db 11733 CTTGGGGGCCCTGGCAGGTTGTGTATGCCCTTGTGTGGACAGGGAA - CAGGAGGAGGAGC 11791
|||||
QY 1681 agacactcgggaggtgggagccagtggtgtctgtgatatgtggcaggtctcagttg 1740
|||||
Db 11792 AGACNACTGGGAGAGTGGGAGGCCAGTGTGTCTGTGGATATGTGGCCAGGTTCAAGTG 11851
|||||
QY 1741 gaagctgaaggatgagcagaccttaggctcaggaagagggtgcctggaagtgggggca 1800
|||||
Db 11852 GAAGCTGAAGGATGAGCAGACCTTAGGCTCAGGAAGAGGGCTGCGCTGGAAGTGGGGCA 11911
|||||
QY 1801 tcactcactgaccagaaaaagggaaactggcagtgccagggctggatggggccctgcattgag 1860
|||||
Db 11912 TCATCACTGACCAGAAAAGGAAACTGGCAGTGCACAGGCTGGATGGGCGCTGCATTGAG 11971
|||||
QY 1861 cttgaaaaaactataatagaattggttaccatttttattttattttattttattttattt 1920
|||||
Db 11972 CTTGAAAAAAACTAATAAGAAATGGTTACCATTTTCAATTTTATTTATTTATTTATTT 12031
|||||
QY 1921 tactttttgagatagagctcactcctctgctaaggtcggagtcggtgggtgctatctc 1980
|||||
Db 12032 TACTTTTTTGTAGATAGAGTCTCAC TCCCTTGCTAAGG - TGGAGTGGGTGGTGTATCTC 12090
|||||
QY 1981 agtcaactgcaacctctgctccccaggatcaagtgatctctcagcctcagcctccccagg 2040
|||||
Db 12091 AGCTCACTCAACCTCTGCTCTCCAGGATCAAGTGATTTCTCCAGGCTCAGCCTCCCCAG 12150
|||||
QY 2041 tagctgggattacaagcatgcccaccatgcctcgataatttttatttttagttgaga 2100
|||||
Db 12151 TAGCTGGGATTAACAGCATGCACCACTGCCCTGGATAATTTTGTATTATTTTAGTTGAGA 12210
|||||
QY 2101 cgggggtttcacaggttggccagactgggtctcgaactctctgacctcaggtgactgcctg 2160
|||||
Db 12211 CGGGGTTTCCAGAGTTGGCCAGACTGGTCTCGAACTTCTGACCTCAGGTGATCTGCCTG 12270
|||||
QY 2161 cctcgccctcccaaggtgtcgaattacaagatgtgagccactgtccctggcctggttacc 2220
|||||

Db 12271 CCTCGCCCTCCCAAGTCTGGAATACAGATGTGAGCCACTGTCCCTGGCGTGTACC 12330
Qy 2221 cacattttaaatgagtgatttcaccccttttatgtgatttaacagctgttttttttt 2280
Db 12331 CACATTTTAAATGAGTGATTTACCCCTTTATGTGGATTTACAGCTTG--TTTTTTT 12388
Qy 2281 ttttttgagacaaagtctggctctgtcccccaggctgtagtgcagtaataatgcacgcagc 2340
Db 12389 TTTTGTGAGACAAAGTCTGTTCTCTCACCCAGGCTGGAGTGCAGTAATGCAATCTCAGC 12448
Qy 2341 tcaactgaacctagctcctctgggttcaagcaattctcctgcctcagccaactgagtagc 2400
Db 12449 TCACGTCAACCTTAGCTCTCTGGTTCAGCAATTTCTCTGCTCAGCCACCTGAGTAG- 12507
Qy 2401 ctggggttacagcatgcacacacagccagcgaactaatttttttttttttttttttttt 2460
Db 12508 CTGGGATTACAGGCATGACCAACCCAGCCAGGCTAATTTTGTATTTTGTAGTAGATG 12567
Qy 2461 ggggttcgcattgttggccaggctggtctcgaactcctgcactcagctcagtgatccgcccgc 2520
Db 12568 GGGTTTCGCCATGTTGGCCAGCGTGTCTCGAATCTCTGACCTCAGGTGATCCGCCGCC 12627
Qy 2521 tbgcctcccaaaagtctaggtattacaggttggaacacctgtccagcctgtggtctatc 2580
Db 12628 TTGGCTCTCCAAAGTGTAGGATTACAGGTGGGAACCACTCGCCAGCCTGTGGCTATC 12687
Qy 2581 gtttaaacactgggaagcctgcagccccagccagccagtagtagctgagctgaagctt 2640
Db 12688 GTTTAAACACTGGGAAGGCTCTGACGCCCCAGCCGACAGTTAGTGCAGCTGACGAGTT 12747
Qy 2641 ccagtgccaggtagcaggtatctcaccactactcactgctgctgctatcttctgtcatagt 2700
Db 12748 CCCAGTGCAGGTAGCAGGATCTCCACCCTACTCATGGCTGATCTCTGTCTATAGT 12807
Qy 2701 gaagtctctggacagaccttcattgttgatctctgttcccccagagtggtggcaat 2760
Db 12808 GAAGTCTCTGGACAGACCTTCACTGTTATGGATCTCTGTTCCCGACAGAGTGGTGGCAAT 12867
Qy 2761 gaatggagtgacaagctcactgggtgttaggggagcagagggccagagtcagagtagata 2820
Db 12868 GAATGGAGTGGACAGCTCACTGGGTGTAGGGGGCAGAGGGCCCAAGTCCAGAGTGTA 12927
Qy 2821 cccccagagtggttccagcagagagcttgcagaggtatctggatggagcagagaggtgg 2880
Db 12928 CCCCCAGAGTGGTCCAGCAGAGCTTCCCGAGGGATCTGGGATGGAGCAGAGGGTGG 12987
Qy 2881 agggggagaccagagaagggggaactgtggtccctgggtgggtctggagtgccctggag 2940
Db 12988 AGGGAGGAGACCCAGAAGAGGGGGAACCTGTGGGCCCTGGGTGGGTCTGGAGTGCCTGGAG 13047
Qy 2941 gaagcccagggcagagagagagagatggatgggtggcagccccagagctggggccgacc 3000
Db 13048 GAAGCCAGGCCAGAGAGAGAGATGGATGGGTGGCGAGCCCCAGGCTGGGCCGACC 13107
Qy 3001 tcacactgtctctgtgcccctgcggtgaccaggtgagccatacgtgcctcaagggc 3060
Db 13108 TCACACTGTGCTCTGTGCCCTGCGCTGGACAGGTGAGCCATACGTGCGCATCAAGGCC 13167
Qy 3061 tcaactgtctgtaggggagacaggtgtccctgtctcgaggggtgaagctgttgagtagtaatt 3120
Db 13168 TACACTGTCTGTGGAGGGGACAGAGTGTCCCTGCTCGAGGGGTGAAGCTGTTGAGGTCAAT 13227
Qy 3121 cacaagctctgagcgtctgtggtcatcaggtagggggccctctccatccagagaca 3180
Db 13228 CACAGCTCTCGAGCGCTGTGGTGTATCAGTGTAGGAGGGCCCTCTCCATCCAGAGCA 13287
Qy 3181 cccatctgagtcagccccagcagcaggggtgttttaggatctgggtgactgttgcct 3240
Db 13288 CCCATCTGAGTCAGCCCCAGCCAGGAGCGCGTGTAGGGAATCTGGGTGACTTGTCCCT 13347
Qy 3241 gggactctgggttaagccactgcctctctctgtgggttagtttccatctcagtagcagggag 3300
Db 13348 GGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAG 13407

Qy 3301 ggaatgagccacccttgctgtcttgggagatccaatgtccttgccttccaagtgggtgcat 3360
Db 13408 GAATGAGCCACCCCTTGCTGTCTGTGGGATCCAATGTCCTTGTCCAAGTGGTGAT 13467
Qy 3361 t 3361
Db 13468 T 13468
RESULT 7
HS47P43S02 3348 bp DNA PRI 07-JAN-1998
LOCUS Homo sapiens p47-phox (NCF1) pseudogene, clone P43, exons 6-8.
DEFINITION U60971
ACCESSION U60971 GI:2754715
VERSION U60971.1
KEYWORDS 2 of 3
SEGMENT human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
JOURNAL 2 (bases 1 to 3348)
MEDLINE Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
REFERENCE Christensen,B., Curnutte,J.T. and Gorlach,A.
AUTHORS Genomic structure and the identification of multiple polymorphisms
in the p47-phox gene
unpublished
JOURNAL 3 (bases 1 to 3348)
REFERENCE Gorlach,A., Lee,P.L., Roesler,J., Chanock,S.J. and Curnutte,J.T.
AUTHORS Direct Submission
TITLE Submitted (17-JUN-1996) Molecular and Experimental Medicine, The
JOURNAL Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
Location/Qualifiers
1..3348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P1 clone P43"
/chromosome="7"
/map="7q11.23"
/note="Human genomic P1 clone P43 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2"
507..629
/gene="NCF1"
/pseudo
/number=6
1093..1200
/gene="NCF1"
/pseudo
/number=7
2749..2866
/gene="NCF1"
/pseudo
/number=8
715 a 854 c 1003 g 776 t
BASE COUNT
ORIGIN
Query Match 89.1%; Score 2993.2; DB 9; Length 3348;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 3052; Conservative 0; Mismatches 23; Indels 5; Gaps 4;
Qy 282 aaaaaaaaaaaagtgggaaagggccaggtgcagtgctccagcgcctgtaatcc 341
|||||

Qy 2502 ctccagtgatccgcccgccttggccctcccaagtctagattacagatggaaacacct 2561
|||||
Db 2216 ctccagtgatccgcccgccttggccctcccaagtctagattacagatggaaacacct 2275
|||||
Qy 2562 tgcccagcctgtggctatctgtttaaacactgggaaggcctcagccccagcgcagcagt 2621
|||||
Db 2276 cgcccagcctgtggctatctgtttaaacactgggaaggcctcagccccagcgcagcagt 2335
|||||
Qy 2622 tagctcagctgagcagttcccaagtcccaagttagacgtagatctcccaaccactactatg 2681
|||||
Db 2336 tagctcagctgagcagttcccaagtcccaagttagacgtagatctcccaaccactactatg 2395
|||||
Qy 2682 gctgctctctgctatagtaagtcttgacacacttccatctgtagatctctgct 2741
|||||
Db 2396 gctgctctctgctatagtaagtcttgacacacttccatctgtagatctctgct 2455
|||||
Qy 2742 ccccagagtgggtgcaatgaatggagtgagacacacttccatctgtagatctctgct 2801
|||||
Db 2456 ccccagagtgggtgcaatgaatggagtgagacacacttccatctgtagatctctgct 2515
|||||
Qy 2802 gcccagaatccagagtgtaacccccagagtgggtgcccagcagagagcttccagaggatctg 2861
|||||
Db 2516 gcccagaatccagagtgtaacccccagagtgggtgcccagcagagagcttccagaggatctg 2575
|||||
Qy 2862 gcatgagcagaggggtgagggagagacaccagaggggggaactgtggccctgggt 2921
|||||
Db 2576 gcatgagcagaggggtgagggagagacaccagaggggggaactgtggccctgggt 2635
|||||
Qy 2922 gggctgagtgcttggaggagcccgagcagagagagagagagagagagagagagagagag 2981
|||||
Db 2636 gggctgagtgcttggaggagcccgagcagagagagagagagagagagagagagagagag 2695
|||||
Qy 2982 gcccaggtgggcccagactcaactgtgctgtgcccagcagagagagagagagagagagag 3041
|||||
Db 2696 gcccaggtgggcccagactcaactgtgctgtgcccagcagagagagagagagagagagag 2755
|||||
Qy 3042 atacgtcgccatcaaggcctacactgtgtgagggggagcagaggtgtccctgctcagagg 3101
|||||
Db 2756 atacgtcgccatcaaggcctacactgtgtgagggggagcagaggtgtccctgctcagagg 2815
|||||
Qy 3102 tgaagctgtgtgaggttaattcaaacagctcctggagcagagagagagagagagagagag 3161
|||||
Db 2816 tgaagctgtgtgaggttaattcaaacagctcctggagcagagagagagagagagagagag 2875
|||||
Qy 3162 cccctctccatccagacacccatctgagtcagccccagcagagcaggggtgttttaggga 3221
|||||
Db 2876 cccctctccatccagacacccatctgagtcagccccagcagagcaggggtgttttaggga 2935
|||||
Qy 3222 tctgggtgacttgccttgggactctgggttaagccactgcccctctcttgggcttagttt 3281
|||||
Db 2936 tctgggtgacttgccttgggactctgggttaagccactgcccctctcttgggcttagttt 2995
|||||
Qy 3282 ccattcagtaagcagggagagatagaccaccccttgcctgtcttcttggggtatccaatgtc 3341
|||||
Db 2996 ccattcagtaagcagggagagatagaccaccccttgcctgtcttcttggggtatccaatgtc 3055
|||||
Qy 3342 cttgtcccaagtgggtgcatt 3361
|||||
Db 3056 cttgtcccaagtgggtgcatt 3075
|||||

RESULT 8
HS47P40S02
LOCUS HS47P40S02 3350 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox (NCFL) pseudogene, clone P40, exons 6-8.
ACCESSION U61239
VERSION U61239.1 GI:2754722
KEYWORDS
SEGMENT 2 of 3
SOURCE Homo sapiens
ORGANISM human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3350)
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
2 (bases 1 to 3350)
Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
Characterization of the genomic structure of the p47-phox gene
Unpublished
3 (bases 1 to 3350)
Gorlach,A., Lee,P.L., Roesler,J., Chanock,S.J. and Curnutte,J.T.
Direct Submission
Submitted (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
Location/Qualifiers
1..3350
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P40"
/chromosome="7"
/map="7q11.23"
/note="Human genomic P1 clone P40 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2"
507..629
/gene="NCFL"
/pseudo
/number=6
1093..1200
/gene="NCFL"
/pseudo
/number=7
2749..2866
/gene="NCFL"
/pseudo
/number=8
BASE COUNT 711 a 856 c 1008 g 775 t
ORIGIN
Query Match 89.0%; Score 2990; DB 9; Length 3350;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 3050; Conservative 0; Mismatches 25; Indels 5; Gaps 4;
Qy 282 aaaaaaaaaaaaaagttggaaaagccaggtgcagtggtccacgcgctgtaatcc 341
|||||
Db 1 aaaaaaaaaaaaaagttggaaaagccaggtgcagtggtccacgcgctgtaatcc 60
|||||
Qy 342 caacacttaagagctaggtggagagaaatctttgagccccagagagttcgagaccagct 401
|||||
Db 61 CAACACTTAAAGAGCTGAGGTGGAGAAATCTTTAGCCCCAGGAGTTCGAGCAGGCT 120
|||||
Qy 402 gggcattgtcccaagaccttctttacaaaaatagccgggtgtgtgggtacacgtct 461
|||||
Db 121 GGGCATTGTCCCAAGACCTTGTCTTTACCAGAAATAGCCGGGTCTGTGGCATACGTC 180
|||||
Qy 462 gtggctccagctattcgggagagctgagcagggagattgttgagcctaggagctaggg 521
|||||
Db 181 GTGGTCCCGAGCTATTTCGGGAGGCTGAGGCAGGGAGAGATTGCTTGAAGCTAGGAGT 240
|||||
Qy 522 ctgtagtagctgtgatacagtcactgtactactagcctgggcaacagagagagctctgt 581
|||||
Db 241 CTGTAGTAGCTGTGATCAGCTCACTGTACTACTGCTGCGGCAACAGAGCAAGACTCTGT 300
|||||
Qy 582 ctccaaaaaagaaaaaagttggaaaagggctcactaaccttcacagatgagaaacaaaga 641
|||||
Db 301 CTCCCAAAAAAATAAAGTTGGGAAAGGCTCACTAACTTCATCAGATGAGAACAAAGA 360
|||||
Qy 642 catgtttgaagtgtgagccgaagcctggagaaacgtatgcgccccaggaatgcaggggca 701
|||||

|||||
Db 361 CATGTTTAAAGTGTGAGCCCAAGCCTGGAGAACCTATGCGCCAGGAAATGCAGGGCA 420
Qy 702 gcagagactcaagatgcagcgccctgtctgtgagccagatggccctgcaatgccac 761
Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTTCTGAGGCCAGATGGCCCTGCAATGCCAC 480
Qy 762 tcaacctgcctcctctctgcccagacatcaaccggcccctcatcctgcagagctaccg 821
Db 481 TCACCTGCCCTCCCTCTGTGCCCCAGAGACATCACCGGGCCCCATCATCTCTGCAGAGCTACCG 540
Qy 822 cgcattgcagactcacagaagacctcgggctccagatggctctgtccacggggagct 881
Db 541 GGCATATGCCAACTACGAGAAGACTCGGGCTCCGAGATGGCTCTGTCCACGGGGGAGCT 600
Qy 882 ggtggagctgtgagaaagcagagcgggtcagacctcccaccttacggggtccctccc 941
Db 601 GGTGAGGTCTGGAGAGACGGAGAGGGGTTCAGACCTCCCACTTACGGGACTCTTCC 660
Qy 942 cctggtgtcaggaaccacacagcccaagccccctgccaaggtctcagcagcctggcccc 1001
Db 661 CTTGTTGCTCAGGAACCCACAGCCACAAAGCCCCCTGCCAA-GCTCAGGACAGCTGGCCCC 719
Qy 1002 tggagagactcagctctgttaggggcccctaaatgtcctcccacactgtgggtcgccctt- 1061
Db 720 TGGGAGGACTCCGGCTCTGTTAGGGGGCCCTAAATGTCTCTCCCACTGTGGGTCTGCGCTT 779
Qy 1062 ctctcttagtgcacacctgtgtgtgtgtggtcgtgtgcatgtgcaaggcggggcggg 1121
Db 780 CTCCTTGTAGTGTGACCCCTGTGTGTGTGTGGGTCTGTGTGATGCGAGCGGGCGGG 839
Qy 1122 gcatgtctgcgttctgtctgtggtggtatgggatagggacctgtgttcaattatgaagtgggct 1181
Db 840 GCATGCTCTGCGCTGTCTCTCGGATGGGTATGGGACCTCTGTTCATTTATGAAGTGGGCT 899
Qy 1182 cagagctgtgattctgtgagcatgtgtgcatgcatgcatgtgacctcaattgtccagtg 1241
Db 900 CAGAGCTGTGTATCTGTGTGAGCATGTGTGCATGTGCATGTGCATGTGCATTTCCAGTGTG 959
Qy 1242 gtgaaggtgacatttcaaaatctgagcatgtgacatcagtgctgtgtgtcctgtgtcc 1301
Db 960 GTGAAGGTGACATTTCCAAATCTCAGCAATTCAGATCATAGTGTGTGTGTGTGTGTGTGTGT 1019
Qy 1302 tcacatccctgtatgggtgcaggagccgtgggcccctgcctccctcagtcacattcccga 1361
Db 1020 TCACCATCCCTGATGGCTGCAGGAGCGCTGGGCCCTGCCCCCTCAGTCACATTTCCCGCA 1079
Qy 1362 cctctggcaaggtgtgtgtgtctctgaagtgaaagcaaaagcgaggctggatccccagcat 1421
Db 1080 CCTCTGGCACAGGTGTGTGTGTCTGTCAAGATGAAGCAAAAGCGAGGCTTGGATCCAGCAT 1139
Qy 1422 ccttccctgagccctggacagctcctgacgagacggaagaccctgagcccacactatcag 1481
Db 1140 CTTTCTCTCAGCCCCCTGGACAGTCTTGCAGAGAGCGGAAGACCCCTGAGCCCCAACTATGCAG 1199
Qy 1482 gtgcccctgcctccagggctgtagggtgtgtggagaaagggcagggagggctcaggg 1541
Db 1200 GTGCCCCCTGCCCCCGAGGGCTGTAGGGGTGTGGGAGAAAGGGGACGGCAGGGCTCAGGG 1259
Qy 1542 atattgagtgactccttggagctgggctgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1601
Db 1260 ATATTGAGTGACTCTTGTGGAGTCTGGGCTGTGGCTGTGGCTTGGCAGAAAGTCAAGGGCT 1319
Qy 1602 aagatcctatcgctctgtgcttggggccctggcaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
Db 1320 AAGATCTCATCTCTGTGGCTTGGGGCCCCCTGGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1379
Qy 1662 gggaaacaggaggagagacagactcgggagagtgaggagggcaggtgtgtgtgtgtgtgtgt 1721
Db 1380 GGGAA-CAGGAGGAGGAGACACAGACATCGGGAGAGTGGGAGGCGGCAGTGTGTGTGTGTGTGT 1438
Qy 1722 atgtggccaggttcaagtgggaagctgaaggatgagcagaccttaggctcaggaaggagg 1781
|||||

Db 1439 ATGTGGCCAGGTTTCACTGGGAAGCTGAAGGATGACAGACCTTAGGCTCAGAAAGAGGG 1498
Qy 1782 ctgcttgaagtgggggcaatcatcactgaccagaagggaaaactggcagtgccaggct 1841
Db 1499 CTGCTGGAAGTGGGGCATCATCACTCACCAGAAAGGAAAACGTGCAGTGCCAGGGCT 1558
Qy 1842 ggaatggggcctgcattgagccttgaaaaaaactaataagaattggttaccatttttatttt 1901
Db 1559 GGATGGGGCCTGCATTTGAGGTGTCAAAAACATTAATAAGAAATGGTTTACCAATTTTCAATTT 1618
Qy 1902 attatttatttatttatttatttatttatttatttatttatttatttatttatttatttattt 1961
Db 1619 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1678
Qy 1962 agtgcgggtgtgtctatctcagctcactgcaacctgtgctcccaggaatacaagtattctc 2021
Db 1679 AGTGGGTGGTGTATCTCAGCTCACTCAACCTCTGCCTCCCAGGATCAAGTGATTCTC 1738
Qy 2022 cagctcagcctccccaggttagctgggatttaaaagca tgcaccacca tgcctgggataatt 2081
Db 1739 CAGCCTCAGCCTCCCGAGGTAGCTGGGATTACAAGCATGCACCACTGCCCTGGGATAATT 1798
Qy 2082 ttgtatttttagttgagacgggtttccaccaggttggccagactggctcgaacctctg 2141
Db 1799 TGTGTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1858
Qy 2142 acctcaggtgactctgcctgcctcgccctcccacaaagtgtggaattacagagatggagcac 2201
Db 1859 ACCTCAGGTGATCTGGCTGGCTTGGGCTCCCAAGTGGTGGAAATTACAGATGTGAGGCAC 1918
Qy 2202 tgcctcctggcctgggttaccacattttaaagtggagtgatttaccaccttttatgtggatt 2261
Db 1919 TGTCTCTGGCTGGTGTACCCACATTTAAATGGAGTGATTTACCCCTTTTATGTGGATT 1978
Qy 2262 tacagcttggttt 2321
Db 1979 TACAGCT--TTTTTTTTTTTTTTTTTTTTTTTTTTTGTGACAAAGTCTGGTCTGTACACCCAGGCTGGAGT 2036
Qy 2322 gcagtaatgcacatctcagctcactgcacaccttagcctctgtggttcaagcaattctcctg 2381
Db 2037 GCAGTAATGCAATCTCAGCTCAGTCAACCTTAGCCTCTCTGGGGTCAAGCAATTTCTCTG 2096
Qy 2382 cctcagccacctgagtagcctgggtttacaggcatgcaccaccacagccagggttaattttt 2441
Db 2097 CCTCAGCCACCTGAGTAG-CTGGGATTACAGGCATGCACACACACAGCCAGGCTAATTTTT 2155
Qy 2442 tgtatttttagtagagatgggggttttcgccaatgttggccaggtgtgtcgcgaactcctgac 2501
Db 2156 TGTATTTTATTTAGTAGAGATGGGGTTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCTCTGAC 2215
Qy 2502 ctccaggtgacccgccttggcctcccacaaagtgtcaggtatcaggtgggaaaccact 2561
Db 2216 CTCAGGTGATTCGGCCGCTTGGCCTCCCAAGTGTCTAGGATTACAGGTGGGAACCACT 2275
Qy 2562 tggccagcctgtgggtatcggtttaaacactgggaaggcctgcagccccccagggccagact 2621
Db 2276 GCGCCAGCCTGTGGGTATCTGTTTAAACACTTGGGAAGGCTGCAGCCCCCAGGCCACAGT 2335
Qy 2622 tagctgcagctgcagcagttcccagtgccaggtagacgagtagtccaccacctactcatg 2681
Db 2336 TAGTGCAGCTGAGCAGTTCCCACTGCGAGGTAGACGATGCTCCACCCACCTACTCATG 2395
Qy 2682 gctgatcctctgcatagtgaggtgtctgagcagacctcatcgttatgggactcctggt 2741
Db 2396 GCTGATCTCTTGTCTATAGTGAAGTGTCTGGACAGACCTTCATCGTTATGGGATCTCTGTT 2455
Qy 2742 cccagagtggtgtggaatgaaatgggagtggaacaagctcaccctgggtgttagggggcagag 2801
Db 2456 CCCCAGAGTGGGTGGCAATGAATGGGAGTGGCAAGCTCACCTGGGTGTAGGGGGCAGAG 2515
Qy 2802 ggcggaagtcagaggtgtacccccagagtggtggcagcagagagcttcccagggagatctg 2861
Db 2516 GGGCGAAGTCCAGAGTGTACCCCCAGAGTGGGTGGCCAGAGGAGTTCAGGAGGATCTGTG 2575
|||||

Qy	2862	ggatgacagagaggttgagaggaggaagaccagaaagaggggaactgtggccctgggt	2921
Db	2576	GGATGACAGAGAGGTGGAGGAGAGAGACCCAGAAAGGGGGAACCTGGGCCCTGGGT	2635
Qy	2922	gggtctgagtcctgaggaagccagagcgagagagagaagatggatgggtggcga	2981
Db	2636	GGGTCTGAGTGCCTTGGAGGAAGCCAGAGCGCAGAGAGAGAGATGGGATGGTGGGA	2695
Qy	2982	gccccaggtgggcccacactgtctctgtgtgccccctgcccgtggaccaggtgagcc	3041
Db	2696	GCCCCAGGCTGGGCCGACCTCACACTGTCTGTGTGCCCTGCCGTGACACAGGTGAGCC	2755
Qy	3042	atacttcgcatcaaggcctacactgtgtgaggggagagaggtgtccctgctcgaagg	3101
Db	2736	ATACCTTCGCCATCAAGGCTACATGCTGTGTGAGGGGACGAGGTGTCCTGCTCGAGGG	2815
Qy	3102	tgaactgttgaggtaattcaagaactcctggagcgctgggtgggtcatcaggtagaggg	3161
Db	2816	TGAAGCTGTGAGGTAAATCACAAGCTCTTGACGGCTGTGTGGTTCATCAGTAGGAGGG	2875
Qy	3162	cccctctcattccagagacccattcagtcagccccagggagaggggtgttttaggga	3221
Db	2876	CCCCCTCTCATCCAGAGACCCATCTGAGTCAGCCCCCAGCAGCGCGTGTTAGGGA	2935
Qy	3222	tctgggtgactgtccctggagactctggtaagccactgcccctctctggtcttagttt	3281
Db	2936	TCTGGGTGACTTGTCCCTGGGACTCTGGGTAAAGCCACTGCCCTCTCTGGGCTTAGTTT	2995
Qy	3282	ccatctcagtagcagggagtagccaccccttgctgtcttctgttgggagatccaatgtc	3341
Db	2996	CCATCTCAGTAGCAGGAGGATGAGCCACCCTTGCTGTCTTGTGGGATCCAAATGTC	3055
Qy	3342	cttgtccaaagtggtgcatt	3361
Db	3056	CTTGTCCAAGTGGGTGCATT	3075
RESULT	9		
LOCUS	HS47PHOX02		
DEFINITION	Homo sapiens p47-phox (NCF1) gene, exons 6-8.	PRI	07-JAN-1998
ACCESSION	U57834		
VERSION	U57834.1	GI:2754710	
KEYWORDS	2 of 3		
SEGMENT	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 3349)		
AUTHORS	Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B., Green,E.D., Chanock,S.J. and Currutite,J.T.		
TITLE	A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease		
JOURNAL	J. Clin. Invest. 100 (8), 1907-1918 (1997)		
MEDLINE	97474758		
AUTHORS	2 (bases 1 to 3349)		
REFERENCE	Chanock,S.J., Roesler,J., Hopkins,P., Lee,P., Bassett,D.T., Christensen,B., Currutite,J.T. and Gorlach,A.		
TITLE	Genomic Structure and the Identification of Multiple Polymorphisms in the p47-phox Gene		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 3349)		
AUTHORS	Gorlach,A., Roesler,J., Christensen,B., Chanock,S.J. and Currutite,J.T.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-MAY-1996) P. Lee, Molecular and Experimental Medicine, Scripps Research Institute, 10550 North Torrey Pines Rd, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1..3349		
	/organism="Homo sapiens"		

exon	/db_xref="taxon:9606"	
	/chromosome="7"	
	/clone="PI Clone 42"	
	/map="7q11.23"	
	507..629	
exon	/gene="NCF1"	
	/number=6	
	1092..1199	
exon	/gene="NCF1"	
	/number=7	
	2748..2865	
	/gene="NCF1"	
	/number=8	
BASE COUNT	714 a 860 c 1002 g 773 t	
ORIGIN		
Query Match	88.8%	Score 2984.4; DB 9; Length 3349;
Best Local Similarity	99.1%	Pred. No. 0;
Matches 3053; Conservative	0; Mismatches	21; Indels
6; Gaps	5;	

Qy	282	aaaaaaaaaaaaaaaaaattgggaaaaagccaggtgcagtggtccacgcctgtaattcc	341
Db	1	AAAAAAAAAAAAAAAAAGTTGGGAAAAAGGCAGGTGCAGTGGCTCCACGGCTGTATTC	60
Qy	342	caacactttaagagagctgaggtgggagaaatcctttgagccagagagtttcgagaccagct	401
Db	61	CAACACTTAAAGAGCTCAGGTGGGAGAAATCTTTGAGCCCCAGAGATTCGAGACCAGCT	120
Qy	402	gggcaattgcccagaccttctttacaaaaaattagccgggtgtgtgagacacagct	461
Db	121	GGGCATTGTCCCAAGACCTTGTCTTTACAGAAATTAGCCGGGTGTGGTGGCATACGCT	180
Qy	462	gtgtccagctatttcggagagctgagccagagagattgcttagcctagagcctagagctagg	521
Db	181	ATGGTCCAGCTATTTCGGGAGGCTGAGGAGGAGATTGCTTGAGCCTTAGGAGTCCAGGG	240
Qy	522	ctgtagtgcgtgtatcacgtcactgtactctagctctggcagagagagcaagactctgt	581
Db	241	CTGTAGTCAGCTGTGTATCAGTCACGTCCTACTGTACTTACCTGGGCAACAGAGCAAGCTGT	300
Qy	582	ctccaaaaaagaaataaagttggaaaggctcactaactcattcattcagatgagaaacaga	641
Db	301	CTCCAAAAAGAAATAAAGTTGGGAAAGGCTCACTAATCTTCATCAGATGAGAAAGGA	360
Qy	642	catgtttgaagtgtgagcccaagcctgagaaagcctatgcgccagagaaatgcaggga	701
Db	361	CATGTTTGAAGTGTGAGCGCCGAAGCCTGGAGAACCTATGCGCCCGAGGAATGCAAGGCA	420
Qy	702	gcagagactcaagatgcagagcgctgttctgagggccagagatggccctgcaatgccac	761
Db	421	GCAGAGACTCAAGATGCCAGCGCTGTCTGTGAGGCCCGAGATGGGCCCTGCAATGCCAC	480
Qy	762	taacctgcctccctcttctgcccagacataccagggcccccattccttcgagagctagc	821
Db	481	TCACCTGCCTCCCTCTTTCGCCAGAGATCACCGGCGCCCATCATCTCTGCAGAGTACCG	540
Qy	822	cgcattccagctacagagagacctcggtccgagatggctctgtccacggggagcgt	881
Db	541	CCCCATTGCCAACTACGAGAGAGACCTCGGGCTCCGAGATGGCTCTGTCCACGGGGAGCT	600
Qy	882	ggtgaggtcgtggagaagagcgagcggtcagacctcccacaccttaagggggctcttcc	941
Db	601	GGTGCAGGTCTAGAGAGAGCGGAGAGCGGTTCAGACCTCCACACCTTACGGGCTCTTCC	660
Qy	942	cctggtctcaggaaacccacagccagccccctgcacaggctcagcagcctggcccc	1001
Db	661	CCTGTGTCTCAGAAACCCACAGCCAAAGCCGCCCTTGCCAA-GCTCAGGAGCCTGGCCCC	719
Qy	1002	tggagagactccagctctgttagggccctaaatctccccacactgtggtgcctt	1061
Db	720	TGGGAGGACTCCGGCTCTGTAGGGGGCCCTAAATGTCTCTCCACACTGTGGGTGCCTT	779

Db 2935 TCTGGGGTACTTGTCCCTGGGACTCTGGTAAAGCCACTGGCCCTCTCTGGCTTAGTTT 2994
Qy 3282 ccattctcagtagcaggaggagatgagccacccttgctgtctgttggggatccaatgtc 3341
Db 2995 CCATCTCAGTAGCAGGGAGGAATGAGCCACCCCTTGGCTGTCTGTGGGGATCCAAATGTC 3054
Qy 3342 ettgtccaagtgggtgcatt 3361
Db 3055 CTTGTCCAAGTGGGTGCATT 3074

RESULT 10
LOCUS HS47P41S03 3200 bp DNA PRI 07-JAN-1998
DEFINITION Homo sapiens p47-phox pseudogene, clone P41, exons 6-8.
ACCESSION U61243
VERSION U61243.1 GI:2754727
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 3200)
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing
p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758
REFERENCE 2 (bases 1 to 3200)
Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
TITLE Characterization of the genomic structure of the p47-phox gene
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3200)
Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
FEATURES
source
1..3200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P41"
/map="7q11.23"
/note="Human genomic p1 clone P41 is a p47-phox pseudogene
which is defined by the presence of a GT deletion at the
beginning of exon 2."
359..480
exon
/pseudo
/number=6
944..1051
exon
/pseudo
/number=7
2600..2718
exon
/pseudo
/number=8
BASE COUNT 662 a 823 c 971 g 744 t
ORIGIN

Query Match 84.7%; Score 2845.8; DB 9; Length 3200;
Best Local Similarity 99.1%; Pred. NO. 0;
Matches 2904; Conservative 0; Mismatches 22; Indels 5; Gaps 4;

Qy 431 aaaaattagccgggtggtggcatacgtctgtgtgtccagcatttcggggagcgtgagcc 490
Db 2 AGAAATTAGCCGGGTGTGTGGCATACTGCTGTGTGCCAGCTATTTCGGGAGGCTGAGGC 61

Db	1141		TGGTGTCTGGCTTGGCAGAAAAAGTCAGGGCTAAGATCTCATGCTCTGCTTGGGGGCC	1200
Qy	1631		ctggcagggttgtagtgcctcttgctgtgcacagggaacacaggagagagagacagcactcg	1690
Db	1201		CTGGCAGGTTGTGATGCCCTTGCTGTGCACAGGGAA-CAGGAGGAGGAGCAGACGACTCG	1259
Qy	1691		ggagagtgggaggccagtggtgtctgtgtggatatgtggtccagggttcagtggaagctgaag	1750
Db	1260		GGAGAGTGGGAGGCCAGTGGTCTGTGTGGATATGTGGCCAGGTTCACTGGGAAGCTGAAG	1319
Qy	1751		gatgaacagaccttaggtctcaggaagaaggctgacctggaagtgggggcaatacaactga	1810
Db	1320		GATGACGACACCTTAGGCTCAGGAAGGAGGGCTGCCGTGAAGTGGGGCATCATCACTGA	1379
Qy	1811		ccagaagaaggaaaaactggcagtgccacagggtctgagtggggccctgcatlgagcttgaaaaa	1870
Db	1380		CCAGAAAGGGAACACTGGCAGTGCCAGGGCTGGATGGGGCTCGATTCAGCTTGAAGAAA	1439
Qy	1871		actataatagaattgggtaccattttatttattattattattattatttacttttttg	1930
Db	1440		ACTATAATAGAAATTGGTTACCACTTTCATTTTATTATTATTATTATTACTTTTTTG	1499
Qy	1931		agataagagctcaactccttgtctaaagctggagtgaggctgggtgctgtatctcagotcactgc	1990
Db	1500		AGATAGAGTCTCACTCCTTGTCTAAGGCTGGAGTGCGGTGTGTATCTCAGCTCACTGC	1559
Qy	1991		aacctctgctccacaggatccaagtgtattctcagctcagctccacaggtagctgggat	2050
Db	1560		AACCTCTGCTCCAGGATCAAGTGATTTCTCCAGCTCAGCTCCCCAGGTAGCTGGGAT	1619
Qy	2051		tacaacatgcaccacatgcctggataatttttgtatttttagtgagacggggtttca	2110
Db	1620		TACAACATGCACCACTGCCTGGATAAATTGTGTATTTTTAGTTCAGACGGGGTTTCA	1679
Qy	2111		caagggtggccagactggtctogaactctgaactcagctcaggtgtagctgctcgccctc	2170
Db	1680		CCAGGTTGGCCAGACTGFTCTCGAATTTCTCAGCTCAGGTGATCTGCTGCCCTGGGCTC	1739
Qy	2171		cctaaagtctggaaattacagatgtagcactgtccctggcctggttaccacattttaa	2230
Db	1740		CCAAAGTGGTGGNAATTACAGATCTCAGGCACTGTCCCTGGGCTGGTTACCCACATTTAA	1799
Qy	2231		aatggagtgatttcaccccttttatgtgatttacagctgttttttttttttttttttgaga	2290
Db	1800		AATGGAGTGATTTACCCCTTTTATCTGGATTTACAGCT--TTTTTTTTTTTTTTTTTGA	1857
Qy	2291		caaagtctggctctgtcaaccaggctggagtgagatgaatgcaatctcagotcaactgcaac	2350
Db	1858		CAAAGTCTGGTTCTGTACCCAGGCTGGAGTGCAGTAATGCAATCTCAGCTCACTGCAAC	1917
Qy	2351		cttagcctcctgggttcagaacattctcctgctcagccactgagtagccttgagggttac	2410
Db	1918		CTTAGGCTCCTGGGTCAAGCAATTTCTCTCGCTCAGCCACTCAGTAG-CTGGGATTAC	1976
Qy	2411		aggcatgcaccaccagccaggctaattttttatttttagtagagatgggtttctgcgc	2470
Db	1977		AGGCATGCACCAACCGCCAGGCTAAATTTTTTGTATTTTTAGTAGAGATGGGTTTCGCC	2036
Qy	2471		atgttggccaggtggtctcgaactcctgaactcagctcaggtgataccgcgcgctttggcctccc	2530
Db	2037		ATGTTGGCCAGGCTGGTCTCGAACTCTCTGAACTCTCAGTGTGATCCGCCGCCCTTGCCCTCC	2096
Qy	2531		aaagtctcagattacaggtgggaacacaccttgccagctctggctatcgttttaaacac	2590
Db	2097		AAAGTCTTAGATTACAGTGGGAACCACTTCGCCCAACCTGTGGCTATCGTTTAAACAC	2156
Qy	2591		tgggaagggctgcagccccagccgcagtcagtttagctgcagctgagcagttccacagtgcca	2650
Db	2157		TGGGAAGGCTCGAGGCCCCCGACGACAGTTAGCTGCAGCTGAGCAGTTCACAGTGCCA	2216
Qy	2651		ggtagcggatgctccaccaccactactcatggctgtagctcttctgtaagtgagtgctg	2710

Db	2217	GGTAGAGGAGTGGTCCACCCACCTACTCATGCGCTGATCTCTTTGTCTAGTAGGAAGTGTCTG	2277
Qy	2711	gacagaccttcacggttagggatcctctgggtcccccagagtggtggtggaatgaatggaagt	2770
Db	2277	GACAGACCTTCATCGTTATGGGATCTCTGGTCCCCAGAGTGGGTGGCAATGAATGGGAGT	2336
Qy	2771	ggacagctcaacctgggtgttagggggcagagggccgaagtcacagagtgtaccctccacaggt	2830
Db	2337	GGACAAGCTCACTGGGTGTAGGGGCGAGAGGGCCGAAGTCCAGAGTGTATCCCCACAGT	2396
Qy	2831	ggatgccacagagagcttccagaggatctggatggagcagagaggttgagggagagaga	2890
Db	2397	GGGTGCCACAGAGAGCTGCCGAGGGATCTGGATGGAGCAGGAGGGTGGAGGGAGGAGA	2456
Qy	2891	ccagaaagggggaaacttgggccctgggtgggtctggagtgctcgaggaagacccagg	2950
Db	2457	CCCAGAAGAGGGGGAACGTGTGGGCCCTGGGTGGGTCTGGAGTGCCTGGAGGAAGCCCAGG	2516
Qy	2951	cgagagaggaagaagtgggatgggtggcgagccccaggctggggccagacctcaactctgt	3010
Db	2517	CGCAGAGAGGAGAAGATGGGATGGGTGGCGAGCCCCAGGCTGGGCCGACCTCACACTGTG	2576
Qy	3011	ctctgtgccccctgcctgagaccagtgagccaatacgttcgcatcaagagcctacactctgt	3070
Db	2577	CTCTGTGCCCCCTGCCGTGGACCAGGTGAGCCATACGTCCGCATCAAGGCCCTACACTGCTG	2636
Qy	3071	tggagggggacagaggtgtccccctgctcgaggggtgaagctgttgaggttaattcaacaagctcc	3130
Db	2637	TGAGGGGGACGAGGTGTCCCTGCTCGAGGGGTGAAGCTGTTGAGGTAAATTCACAAGCTCC	2696
Qy	3131	tggagagctgtgtgggttcacaggttagggagggccccctctccatccagagaccocatctgag	3190
Db	2697	TGGACGCCTGGTGGGTTCATCAGGTAGGAGGGCCCCCTCTCCATCCAGAGCACCCACTCTGAG	2756
Qy	3191	tcaagccccagcagagcggggttttaaggatctgggtgactgttccctgggagctctgg	3250
Db	2757	TCAGCCCCACGACGAGGAGGGGGTGTTAGGGATCTGGGTGTGACTTGTCTCCCTGGGACTCTGG	2816
Qy	3251	gtaagcaactgccccctctctgggcttagttccatctcagtagcagggagggatlgagccc	3310
Db	2817	GTAAGCCACTGCCCTCTCTGGGCTTAGTTTCCATCTCAGTAGCAGGGAGGATGAGCCC	2876
Qy	3311	acctgtcctgtctgttggggatcccaatgtccttgtccaaagtgggtgcatt	3361
Db	2877	ACCTTTCCTGTCTTGTGGGGATCCAATGTCTCTGTCCAAGTGGGTGCATT	2927
RESULT 11			
HS47L14S02			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SEGMENT			
SOURCE			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomom			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

TITLE Direct Submission
JOURNAL Submitted (18-JUN-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La Jolla, CA 92037, USA

FEATURES Location/Qualifiers

source 1..2899
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="L14"
 /chromosome="7"
 /map="7q11.23"
 /note="Human genomic EMBL3A bacteriophage clone L14 is a p47-pho pseudogene which is defined by the presence of a GI deletion at the beginning of exon 2"
 join(U61224.1:<879..8119,1..2866)
 /gene="NCF1"
 /pseudo
 /exon 507..629
 /gene="NCF1"
 /pseudo
 /number=6
 /exon 1091..1199
 /gene="NCF1"
 /pseudo
 /number=7
 /exon 2749..2866
 /gene="NCF1"
 /pseudo
 /number=8

BASE COUNT 618 a 618 c 882 g 659 t

ORIGIN

Query Match 84.0%; Score 2823.6; DB 9; Length 2899;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 2880; Conservative 0; Mismatches 19; Indels 5; Gaps 4

Qy 282 aaaaaaaaaaaagtgggaaagccaggtgcagtggtccacgcctgaatcc 341
 |||
 Db 1 AAAAAAAAAAAAAAAAAAGTTGGAAAGCCAGGTGCAGTGGCTCCACGCCTGTAATCC 60

Qy 342 caacactttaagggtcgaggtgggagaatcctttgagcccgaggttcgagaccagcct 401
 |||
 Db 61 CAACACATTTAAGAGGTGAGTGGGAGATCCTTTTGGAGCCAGAGGTTGAGAGCAGCCT 120

Qy 402 ggcattgtcccaagacctgtctttacaaaaattagccgggtggtggcatacgtct 461
 |||
 Db 121 GGGCATGTGCTCCCAAGACCTTGCTCTTACCAGAAATATGCCGGGTGTGTGGCATACGCTCT 180

Qy 462 gtggtccagctattcgggaggtcaggcaggagattgcttgagcctaggagctcaggg 521
 |||
 Db 181 GTGGTCCCAACTATTCCGGAGGCTGAGCGAGGACAGATTGCTTGAGCCTAGCAGTCTAGGG 240

Qy 522 ctgtagtgaactgtatcagctcaactgttacttagcctgggcaacagagaagactctgt 581
 |||
 Db 241 CTGTAGTAGCTGTGATCAGCTACTGTACTTAACCTGGGCAACAGAGCAAGACTCTGT 300

Qy 582 ctccaaaaagaataaaagtgggaaagcgtccactaacttcacagatgagaacaaaga 641
 |||
 Db 301 CTCCAAAAGAAATAAAGTTGGAAAGGCTCACTAACTTCATCAGATGAGAACAAGGA 360

Qy 642 catgtttgaagtgtgagggccgaagcctggagaacgctatgcgccaggaaatgcagggca 701
 |||
 Db 361 CATGTTTGAAGTGTGAGGCGGAAGCTTGAGAAACGCTATGCGCCAGGAAATGCAGGGCA 420

Qy 702 qcagagactcaagatccagcgccctgttctgagcccgagatggccctgcaatgccac 761
 |||
 Db 421 GCAGAGACTCAAGATGCCAGCGCCTGTTCTGGAGGCCCAGATGGCCCTGCAATGCCAC 480

Qy 762 tcaccttgccctccctcttgccccagacatacccgcccccatcatccttgacagactaccg 821
 |||
 Db 481 TCACCTTGCCCTCCCTTTGGCCAGACATACCGCGCCCATCATCTTCGACAGACTACCG 540

Db	1619	ATATATTTATTTATTTATTTATTTTACATTTTTCAGATAGAGTCTCACCTCTGCTAAGGCTGG	1678
Qy	1962	agtcggtggtcgtatctcagctcactgcaaacctctgctcccaaggatacaagtatcttc	2021
Db	1679	AGTCGGGTGCTATCTCAGCTCACTGCAACCTCTGCTCCCAAGGATCAAGTGAATTC	1738
Qy	2022	cagctcagcctcccaggtagctgggatacaagcatgaccaccatgctcgtgataatt	2081
Db	1739	CAGCTCAGGCTCCCAAGTACGTGGATTACAAAGCATGCACCAACCATGCGTGGATAA	1798
Qy	2082	tttgtatttttagtgagcgggggttcaccagggtgcccagactggtctcgaactctg	2141
Db	1799	TTTGTATTTTGTGAGACGGGGTTTACCAGGTTGCCACAGCTGGTCTCGAATCTCTG	1858
Qy	2142	acctcaggtgatcgtcctcctcgccctcccgaagtgcctggaaattacagatgtgagccac	2201
Db	1859	ACCTCAGGTGATCGCTGCCCTGGCCCTCCCAAAAGTCTGGAATTCAGATGTGAGCCAC	1918
Qy	2202	tgctcctggcctgggttaccacaattttaaaattggagtgatttcaaccttttatctgga	2261
Db	1919	TGTCCCTGGCGTGGTTTACCCACATTTTAAATTTGGAATGATTTACCTCTTATCTGGAT	1978
Qy	2262	tacagctctgttttttttttttttttgaacaaagctcgtcctgtcaccagcgtcgagt	2321
Db	1979	TACAGCT- -TTTTTTTTTTTTTTTGTAGACAAAGTCTGGTCTGTGTCACCAAGCTGGAGT	2036
Qy	2322	gcagtaatgcgaactcagctcactgcacaccttagcctcctgggttccagcaattctcctg	2381
Db	2037	GCAGTAATGCAATCTCAGCTCACTGCAACCTTAGCCTCTGGGTTCAAGCAATTTCTCTG	2096
Qy	2382	ctcagcaccctgagtagcctgggtttacaggcatgaccaccacgcagcagctaatctt	2441
Db	2097	CCTCAGCCACTGAGTAG-CTGGGATTACAGGCATGCACCAACCAAGCTAAATTTT	2155
Qy	2442	tgtatttttagtagaatgggttttcgcctatgtggccaggtcgtcctgaaactcctgac	2501
Db	2156	TGTATTTTATAGTAGATGGGTTTTCGCCATGTTGGCCAGGCTGGTCTCGAATCTCTGAC	2215
Qy	2502	ctcaggtgaacctccgcctctggcctcccaagtgctaggattacaggtgggaaccacct	2561
Db	2216	CTCAGGTGATCCGCCGCTTGGCTCCCAAGTCTAGGATTACAGGTGGGAACCACT	2275
Qy	2562	tgccagccttggtctatcgtttaaacactgggaaggcctgcagccccagccgacagt	2621
Db	2276	CGCCCAAGCTGTGGCTATCGTTTAAACACTGGGAAGGCTGCAGCCCCAGCGACAGT	2335
Qy	2622	taqctcagctgaagcagttccacagtgccagtgacaggtatcctccaccacctactcatg	2681
Db	2336	TAGCTGCAGCTGAGCATTTCCCAAGTGCCAGGTAGACGATGCTCCACCCACTACTCATG	2395
Qy	2682	gctgactctctgtcatagtgaagtcctggacagacctctcatgtatggatctctggt	2741
Db	2396	GCTGATCTCTGTTCATAGTAGTGTCTGGACAGACCTTCATCTGTTATGGGATCTCTGGT	2455
Qy	2742	ccccagagtgggtgggcaatgaattggagtggaacagctcactgggtgtagggggcagag	2801
Db	2456	CCCCAGAGTGGTGGCAATGAATGGAGTGGACAAGCTCACTGGGTGTAGGGGGCAGAG	2515
Qy	2802	ggccgaagttccagagtgtacccccagatgggtgccacagagacttgcagagggatctg	2861
Db	2516	GGCCGAAGTCCAGAGTGTACCCCAAGTGGGTGGCAGCAGAGGCTTCCCGAGGGATCTG	2575
Qy	2862	ggatggacagagaggtgagggagagaccagaagagggggaactgtgggccctgggt	2921
Db	2576	GGATGAGACAGAGGTTGGAGGGAGGACCCAGAAAGAGGGGAACCTGTGGGCCCTGGT	2635
Qy	2922	gggtctggagtgcttggaagagccagcgccagagagagaaagtgggaagggtggcgga	2981
Db	2636	GGGTCTGGAGTGCCTGGAGGAAGCCAGCGCAGAGAGAGAAATGGGATGGGTGCGGA	2695
Qy	2982	gccccaggtggccagacctcacactgtctctgtgcccctgcccgtggaccaggtgagcc	3041

Db 2696 GCCCAGGCTGGCGGACCTCACA CTGCTGTGTGCTGCCCTGGCGGACCGAGGTCAGCC 2755

Qy 3042 atacgtgcacatcaagcctacactcgtgtgaggggacgaggtgtccctgtcgtcaggg 3101

Db 2756 ATACGTGCCATCAAGCCTACACTGCTGTGGAGGGGACGAGGTGTCCCTGCTCCGAGG 2815

Qy 3102 tgaagctgttgaggtgaattcaagaagctcctcgagcggcgtgtggttcatacaggcaggaagg 3161

Db 2816 TGAAGCTGTGTGAGGTCAATTCAACAAGCTCTCGACGGCTGGTGGCTCATCATCAGTAGGAGG 2875

Qy 3162 cccctctccatccagagaccccat 3185

Db 2876 CCCCTCTCCATCCAGACACCCAT 2899

RESULT 12

AC092405/c

LOCUS

DEFINITION AC092405 184558 bp DNA HTG 04-JUL-2001

SEQUENCE, 6 unordered pieces.

AC092405

VERSION AC092405.1 GI:14595779

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE Olive baboon.

ORGANISM Papio cynocephalus anubis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutharia; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.

REFERENCE 1 (bases 1 to 184558)

Auye, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Warburton, K.D., Zhang, L.-H. and Green, E.D.

NCSC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 184558)

Green, E.D.

Direct Submission

Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717

Government Circle, Gaithersburg, MD 20877, USA

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@hri.nih.gov

----- Project Information

Center project name: ccx

Center clone name: 170F23

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 178116 bases at least Q40

Consensus quality: 179898 bases at least Q30

Consensus quality: 181141 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 184058; sum-of-ctngs

Quality coverage: 10.52x in Q20 bases; agarose-fp

Quality coverage: 9.60x in Q20 bases; sum-of-ctngs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 13828: contig of 13828 bp in length

* 13829 13928: gap of unknown length
* 13929 31257: contig of 17329 bp in length
* 31258 31357: gap of unknown length
* 31358 42795: contig of 11438 bp in length
* 42796 42895: gap of unknown length
* 42896 61486: contig of 18591 bp in length
* 61487 61586: gap of unknown length
* 61587 111639: contig of 50053 bp in length
* 111640 111739: gap of unknown length
* 111740 184558: contig of 72819 bp in length.

FEATURES

source Location/Qualifiers
1. 184558
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone_lib="RP41-170E23"
/clone_lib="RP41"
1. 13828
/note="assembly_fragment"
13929..31257
/note="assembly_fragment"
31358..42795
/note="assembly_fragment"
clone_end:17
vector_side:left
42896..61486
/note="assembly_fragment"
61587..111639
/note="assembly_fragment"
111740..184558
/note="assembly_fragment"
clone_end:SP6
vector_side:right
48162 a 44484 c 44779 g 46622 t 511 others

BASE COUNT
ORIGIN

Query Match 79.4%; Score 2668.8; DB 2; Length 184558;
Best Local Similarity 90.9%; Pred. No. 0;
Matches 3083; Conservative 0; Mismatches 257; Indels 50; Gaps 21;
QY 1 aaaggctcaggcaactgtggctcactcctgttaatccagcaactttgggggcccaggcggt 60
Db 144477 AAAGGCCAGGTGCTGTGGCTTATGCCGTATCTCGGCACCTTTGGGAGGCCAAGGTGGG 144418
QY 61 tgaatcaactaggttcagaggttcagaccagcctgcaccaatatcgtgaacctccatctc 120
Db 144417 TGGATCACTGAGGTGAGGAGTTCAAGACCAGCCCTGACCAATATTTGTAACCTCCATCTC 144358
QY 121 tactaaaaatacaaaaattagcaggtgtggtggcggtgtctttagtcccagctacttg 180
Db 144357 TACTAAAAATACAAAAATTAGCCGGCGTGTGTAGCAGGTGCTTGTAGTCCAGCTACTTG 144298
QY 181 ggagctcaggcaggtgaattacttgacctgggaggtggaggttgcaatgagccaaagt 240
Db 144297 GGAGGCTGAGGTGAGTGAATTCCTTGAACCTGGCAGGTGGAGTTGCAGTAAGCCAAAGT 144238
QY 241 tgcaccactgcactcca-----gtgcagagcagagactccatctc-----aaaaa 286
Db 144237 TGCACCACTGCACTCCAGCTGGCGGCGACAGAGTGAGACTCCGCTCAAAAAAAAAAAAAA 144178
QY 287 aaaaaaataaagtgggaaaaagccaggtgcagtggtccacgcctgtaatcccaaca 346
Db 144177 AAAAAAATAAAGTTGGAAAAAGCCAGGTGCAGTG-----GCCGTGAATCCCAACA 144125
QY 347 ctttaagggtcaggtgggagaaatcctttgagccagcagaggttcgagaccagctgggca 406
Db 144124 CTTTAAGAGGCTGAGGTGGGAGAATTCCTTTAGCCAGGAGTTAAGACCAAGCCCTGGGCA 144065
QY 407 ttgtcccaagacctgtctttcaaaaaaatagccggtgtggtggcatacgtctgtgt 466
Db 144064 TTGTCCCAAGACCTTGTCTTTTCAAAAAAATTAAGCCAGGTATAGTGGCATATGCCCTGCAGT 144005
QY 467 cccagctattcgggaggtgagggagagattgttgagcctaggagcttagggctgta 526

Db 144004 CCCAGCTATTTCGGGAGGCTGAGACAGGGAGATTGCTTGAGCCTAGGAGTCCAGGCTGCA 143945
QY 527 gtgagctgtatcactcactgtactctagctctgggcaagagcaagactctgtctcca 586
Db 143944 GTGAGCTGTGATCATGTCACTGTACTCTTAGCTGGCAACAGCAAGACTCTGTCTCCA 143885
QY 587 aaaaagaaaaataaagtgtggaaagcctcactaaactctcatcatcagatgagacaagaacatgt 646
Db 143884 TAAAGAAACCTTAAAGTTGAGAAAGGCTCACCTAACCTTCACTAGATGAGAACAGGACATGC 143825
QY 647 ttgaagtgtgagcgaagcctgagaaacgctatgccccaggaaatgacagggcagcaga 706
Db 143824 TTGAATTGGGAGGCCAAAGCCTGGAGAACGCTATGCGCCAGGAAATGCA-GGCAGCAGA 143766
QY 707 gactcaagatgcagcgcctgttcttgaggcccagatgagccctctgcaatgcccactcacc 766
Db 143765 GACTCAGGACGCCAAGCGCTGTCTTGGAGGCCAAGATGGGCCCGCAGTGGCCCATCACC 143706
QY 767 ctgcccctcctctgccccagacatcacgcgcccatcatcctgcagacgtaccgcgcga 826
Db 143705 CTGCCCTCCCTC-TGCCCCACACATCACCGGTCCCATCATCTCTGCAGACGTACCCGGCCA 143647
QY 827 ttgcgactacgagaagacctcgggctccgagatggctctctgtcccacggtgggacgtggtgg 886
Db 143646 TTGCCGATTATGAGAAGACCTCAGGCTCTGAGATGGCTTGGCCATGGGGACGCTGTGTG 143587
QY 887 aggtcgtgagaagaagcagagcggtcagacctcccaacttacagggggtctctcccttgg 946
Db 143586 AGGTGCTGGAGAAGACGAGAGCGGTGACAGCTCCCGCCCTACGGGGTCTCTCCCTTAG 143527
QY 947 tgcaggaacccacagcccaagccccctgcgaaggtcagggcagcctgccccctggga 1006
Db 143526 TGCTCAGGAACCCACAGCCCAAGCCCTGCGCAAGGCTCAGACAGCTGGCCCTGGGA 143467
QY 1007 ggaactccagctgtttagggccctaaatgtctctcccccacactgtgggtgcacctctctc 1066
Db 143466 GGACTCTGTTCTTTAGGGGCCCTGAATGTCTCTCCACACTGCGAGTCCGCCCTGTCTC 143407
QY 1067 ttagtgtcacccctgtgtggtgtggcatctgtgagcagccgagccgagggcggtg 1126
Db 143406 TGAGTGTCAACCTTCAGCTGGCTGTGGGCATCTGTGCATGGCAGCCGGGGCGGTGCATG 143347
QY 1127 tctgctgtctctgtctggtggtgtatgggacctgtgttcaattgaagtgggtccagag 1186
Db 143346 TCTGGGTGTTCTGTCTGGTGGGTATGGACCGTCTGCTCATATGAAGTGGGCTCAGAG 143287
QY 1187 ctgtgattctgtgagcagtgtgtgcatgcatgtgacctcattgtccagttgtgtgaa 1246
Db 143286 CTATGATTCTGTGAGCATATGTGCATGTCATGTGACCTCGTTGTCCAGTGAAGTGA 143227
QY 1247 ggtgacatttccaaatctgagcatttgacatcagttgtctgttccctgttctcacc 1306
Db 143226 GGTGACTTTTCCAAATCTGAGCATTTGGACATTTGTGTCTGTGTCTGTCTCTCACC 143167
QY 1307 atccctgtatggctgagggagcgcctgggccccctgcacctcagtcacattcccgcacctc 1366
Db 143166 ATCCCTGATGCTGCAGGAGGCCCTCCGGGTTCCGCCCTCAGTCACATTCGCCGATCTCT 143107
QY 1367 ggcacaggttgggtgtctgttcagatgaagcaagcagagcgtggtatccccagcatcttc 1426
Db 143106 GGCACAGGTTGGTGTGTTCTGTCCAGATGAAGCAAGCAGGCGTGTATCCAGCATCTCTTC 143047
QY 1427 ctgcagccccctgagcagctcctgacgagcgaagacctgagcccaaatatgcaggtgcc 1486
Db 143046 CTCGAGCCCCCTGGACAGTCCCGCCCTGAGAGAACCTGTAGCCCCACTATGACAGGTGCC 142987
QY 1487 cccctccctccgagcgtgttaggggtgtggagaaagggcagggcagggcctcagggatatt 1546
Db 142986 CCCTCCCTCCAGGCCATAGGGGTGTGGGAAAGGGGAGCGGCGGAGGATGTT 142927
QY 1547 gagtgaactgcttggagctctggggtggtgtgctggttggcagaaaaagtcagggtcagaat 1606

Db 142926	GAGTCACTGCTTTGGAGTCTGGACTGCTTCTGCTTGGCGACAAAAGTCAAGGCTGAGAC	142867
Qy 1607	ctcatcgctcgctggctt-ggggggccctggcaggttgtagtcgccttggctctgtgacacagga	1665
Db 142866	CTCATCTTCTCTGCTTGGGGGCCCTTGGCAGGTTGGATATCTTTGGTCCGGACAGGA	142807
Qy 1666	accaggggaggagcacagcactcgggagagtggggagggccagtggtgtctgtggatatgt	1725
Db 142806	ACCAGGAGGAGGAGCAACATGGGGAGGGTGGGAGGTGAGTATATGTCTATGGGTATGT	142747
Qy 1726	ggccagagtccagtgggagcgtgaagatgacagacttaggtctcaggaagaggctcgc	1785
Db 142746	GGTCAGGTCCAGTGGGAAACTGAAGGACGAGCAGACTTAGGTTCAGGAAGAGGGCTGC	142687
Qy 1786	ctggaagtgggggcatcatcactgaccagaaaggaaacctggcagtgccagggctggat	1845
Db 142686	CTGAAGGTGGAGGCGTCATCATTGACCAGAAAGGAAAACTGSCAGTGTGGGCTGGAT	142627
Qy 1846	ggggcctgcattgagct-tgaaaaaaactaatagaattgggttaaccttttatttta-t	1903
Db 142626	GGGGCCTGCATTTGAGCTTCGGAAAAAACTATAATAGAAATGGTTACCATTTATTTTACT	142567
Qy 1904	tatttattttattttact- ----tcttttgatagatgctcactccctgtctaagc	1958
Db 142566	TATTTATTTATTTATCTTTGTGTGTGTGTGAGACAGCGTCTCATCTCCTTGTCTCAGCG	142507
Qy 1959	tggagtcgggtgggtgctatctcagctcactgcgaacctctgc-ctccaggaatcaagtgt	2017
Db 142506	TGGAGTCGAGTGGTGTGATCTCGGTTCACTTGCAATCTCTGCTGTCCGGGTTCAAGTAT	142447
Qy 2018	tctcagcctcagctcccccaggtagctgggattacaagcatgacacacatgccttggat	2077
Db 142446	TCTCCAGCTCAGGCTCCTTGA- GTAGCTGGGATTACAGGCATGTGCCACACACCCAGCT	142388
Qy 2078	aatttttgatttttagtgagacgggggttctaccaggttggccagactgggtctcgaact	2137
Db 142387	AAATTTTGTATTTTAAATAGAGAC- GGGTTTTCACCATTTTGGCCAGGCTGGTCTCGAAT	142329
Qy 2138	ctgaacctcaggtgatctgcctcctcgccctcccaaatgctgggaattacagatgtgag	2197
Db 142328	TCTGACCTCAGGTAAATCCGCTGCCTCAGTCTCCCAAAGGGCTGGGATTTACAGGTATGAG	142269
Qy 2198	ccactgtccctggcctgggttaccacatatttaaaatggagtgatttcaaccttttatgtg	2257
Db 142268	CCACTGCCCTGGCCTGTTTACCCACATTTTAAATGGCGTGATTTCACTTTTATGTG	142209
Qy 2258	gatttaacgc- ----tgcttt	2312
Db 142208	GATTTTACAGCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	142149
Qy 2313	ggctggagtcagtaatgcgaatctcagctcactgcgaaccttagcttccctgggttcaagca	2372
Db 142148	GGCTGGAGTGTAGTAATGCAATCTTGGCTCAGTGCACAACTTCACTTCTTGGGTTCAAGCA	142089
Qy 2373	attctctgcctcagccactgagtcagctggggttacaggcatgcccaccacgcagcg	2432
Db 142088	ATTCTCCACCTCAGCCTCCTGAGTAG- CTGGGATTACAGGCGTGCCACACACGCCAGG	142030
Qy 2433	ctaaatttttgatttttagtagagatgggggtttcgccattgttggccaggtcgggtctcga	2492
Db 142029	CTAA- TTTTGTATTTTATTTTATGATGGGTTTTCACCATGTTTGGACAGGCTGGTCTTGA	141971
Qy 2493	actcctgaacctcaggtgatccgcccgccttggcctcccaagtgtcagattacaggtgg	2552
Db 141970	ACTCCTGACCTCAGTGATCCACCCACTCGGGCCACCCAAAGTTCGGGGGATTACAGCGCG	141911
Qy 2553	gaaccacttgcacagccttgggtctatcgtttaaacactgggaagccttgcagccccccag	2612
Db 141910	GAACCACTTGTCCAGCTGTGGCTATTGTTTAAACACTGGGAAGGCTTGCAGCCCCCAG	141851
Qy 2613	gcgcagacagtttagctgcagctgagcttcccagttccagtgccaggtagacagga	2672
Db 141850	GCCGAC- ----AGGCACAGCCAAAGCAGTTCACCTGTCAGCTGCAGAGGCGCTCCACCCAC	141795

Qy	2673	ctactcagggctgatactcttgatagtgtaagtgctctgacagacacttcacgttata	g99g	2732
Db	141794	CPACTCATGGCTGGTCTCTGTGTGCATATTTAGGTGTCTTGACAGACCTTCATCGTGTGGGA	141735	
Qy	2733	atctctggctccccaagtgaggtggaataaattgagagtgagacaagcttcacctgggtatag	2792	
Db	141734	A--TCTGTGTCCTCCGGAGTGGGTGGCAATGANTGGGAGTGGACAAAGCTCACCTGGGTGTAG	141677	
Qy	2793	ggggcagagggccgaagctccagagtacacccccagagtggggtccacagcagagcgttgcgg	2852	
Db	141676	GGCGCAGAGGGCAGAAAGTCCAGAGTGTGCCCCCAGAGAGGGTGCCACGACGAGCTTCCG	141617	
Qy	2853	aggg--atctggatggagcagagagggatgagggagagagaccacgaagaggggaactgtg	2911	
Db	141616	AGGGAACCCGGATCGAGTAGAGGATGGGGGAGAGAGACCCAAAGA--GGGGCGCTGTG	141558	
Qy	2912	ggccctgggtgggtctggagtgccctggagaaagccacgagcagagagaggaagatggga	2971	
Db	141557	GGCCCTGGGTGGGTCTGGAGTGCCTGGAGAAACCCAGGTGCAGAGAGAGAAGACGGGA	141499	
Qy	2972	tgggtggcagagccccagggctgggcccagacctcacactgtgctctgtgcccctgccgtggac	3031	
Db	141497	TGGGTGGCAGGCCCGGGCTGGCCAGATCTCACACCGTGTCTGTGCCCGCCGCTGGAC	141438	
Qy	3032	caggtagccatacgtcgcacataaagacctacaactgctgtgagggggagcagagtgctcc	3091	
Db	141437	CAGGCGAGGCCGTACGTTGGCATCAAGCGCTACGCTGCTGTGAGAGAGGACGAGGTGTCCC	141378	
Qy	3092	tgtctgaggggtgaagctgttgaggttaattcacaaagctccctggcggctgggtgggtcatca	3151	
Db	141377	TCCTCGAGGGCGAGCTGTTGAGGTCAATTCACAAGCTCTCTGGAGCGGTGGTGGGTCAATCA	141318	
Qy	3152	ggtagagggccctctccatccagagaccccatctgagtcagccccagcagcagagcggg--	3210	
Db	141317	GGTAGGAGGGCTCCGCTCCATCCAGGCGACCCACCCAGAGTCAGCCCCACCCAGGGTGGGA	141258	
Qy	3211	gtgttagggatctgggggtgaactgttccctgggacctctgggtaagcacactgcccctctct	3270	
Db	141257	CTGTTTGGGGATCTGGGGTGACTGTGCCCTGGGACTCTGGGTGAAGGCACATGCCCTCTCTCT	141198	
Qy	3271	gggcttagtttccatctcagtagcagggagggagatgagccacccctgctgctctgttggg	3330	
Db	141197	GGGCTCAGTTTCCATCTCAGTAGCAGGAGCGGTGAGGCCACCCCTGGCTGTCTTGTGGG	141138	
Qy	3331	gatccaatgtcccttgatccaagtgggtgcatt	3360	
Db	141137	GATCCAATGTCTTGTCCAAAGTGCATGCAT	141108	
RESULT	13			
AC027219/c				
LOCUS	AC027219	149830 bp	DNA	HTG
DEFINITION	Homo sapiens clone RP11-729p19, *** SEQUENCING IN PROGRESS ***, 03-MAR-2001			
ACCESSION	AC027219			
VERSION	AC027219.3			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 149830)			
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.			
JOURNAL	Homo sapiens, clone RP11-729P19			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 149830)			
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Balgwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhatat,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,			

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2001 this sequence version replaced gi:11610941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information

Center project name: L8346
Center clone name: 729_P_19

* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1569: contig of 1569 bp in length
* 1570 1669: gap of 100 bp
* 1670 2364: contig of 695 bp in length
* 2365 2464: gap of 100 bp
* 2465 3099: contig of 635 bp in length
* 3100 3199: gap of 100 bp
* 3200 3762: contig of 563 bp in length
* 3763 3862: gap of 100 bp
* 3863 4534: contig of 672 bp in length
* 4535 4634: gap of 100 bp
* 4635 5474: contig of 840 bp in length
* 5475 5574: gap of 100 bp
* 5575 6460: contig of 886 bp in length
* 6461 6560: gap of 100 bp
* 6561 7785: contig of 1225 bp in length
* 7786 7885: gap of 100 bp
* 7886 8735: contig of 850 bp in length
* 8736 8835: gap of 100 bp
* 8836 9864: contig of 1029 bp in length
* 9865 9964: gap of 100 bp
* 9965 10784: contig of 820 bp in length
* 10785 10884: gap of 100 bp
* 10885 11653: contig of 769 bp in length
* 11654 11753: gap of 100 bp
* 11754 12415: contig of 662 bp in length
* 12416 12515: gap of 100 bp
* 12516 13232: contig of 717 bp in length
* 13233 13332: gap of 100 bp
* 13333 14178: contig of 846 bp in length
* 14179 14278: gap of 100 bp
* 14279 15762: contig of 1484 bp in length
* 15763 15862: gap of 100 bp
* 15863 16423: contig of 561 bp in length
* 16424 16523: gap of 100 bp

16524 17467: contig of 944 bp in length
* 17468 17567: gap of 100 bp
* 17568 18533: contig of 966 bp in length
* 18534 18633: gap of 100 bp
* 18634 19882: contig of 1249 bp in length
* 19883 19982: gap of 100 bp
* 19983 21070: contig of 1088 bp in length
* 21071 21170: gap of 100 bp
* 21171 22152: contig of 982 bp in length
* 22153 22252: gap of 100 bp
* 22253 23303: contig of 1051 bp in length
* 23304 23403: gap of 100 bp
* 23404 24381: contig of 978 bp in length
* 24382 24481: gap of 100 bp
* 24482 26183: contig of 1702 bp in length
* 26184 26283: gap of 100 bp
* 26284 27663: contig of 1380 bp in length
* 27664 27763: gap of 100 bp
* 27764 28566: contig of 803 bp in length
* 28567 28666: gap of 100 bp
* 28667 30309: contig of 1643 bp in length
* 30310 30409: gap of 100 bp
* 30410 31434: contig of 1025 bp in length
* 31435 31534: gap of 100 bp
* 31535 32225: contig of 691 bp in length
* 32226 32325: gap of 100 bp
* 32326 33690: contig of 1365 bp in length
* 33691 33790: gap of 100 bp
* 33791 34434: contig of 644 bp in length
* 34435 34534: gap of 100 bp
* 34535 35776: contig of 1242 bp in length
* 35777 35876: gap of 100 bp
* 35877 37225: contig of 1349 bp in length
* 37226 37325: gap of 100 bp
* 37326 38564: contig of 1239 bp in length
* 38565 38664: gap of 100 bp
* 38665 40341: contig of 1677 bp in length
* 40342 40441: gap of 100 bp
* 40442 41665: contig of 1224 bp in length
* 41666 41765: gap of 100 bp
* 41766 42998: contig of 1233 bp in length
* 42999 43098: gap of 100 bp
* 43099 44123: contig of 1025 bp in length
* 44124 44223: gap of 100 bp
* 44224 46278: contig of 2055 bp in length
* 46279 46378: gap of 100 bp
* 46379 47723: contig of 1351 bp in length
* 47730 47829: gap of 100 bp
* 47830 49626: contig of 1797 bp in length
* 49627 49726: gap of 100 bp
* 49727 51471: contig of 1745 bp in length
* 51472 51571: gap of 100 bp
* 51572 52901: contig of 1330 bp in length
* 52902 53001: gap of 100 bp
* 53002 53821: contig of 820 bp in length
* 53822 53921: gap of 100 bp
* 53922 55355: contig of 1434 bp in length
* 55356 55455: gap of 100 bp
* 55456 57801: contig of 2346 bp in length
* 57802 57901: gap of 100 bp
* 57902 59549: contig of 1648 bp in length
* 59550 59649: gap of 100 bp
* 59650 60555: contig of 906 bp in length
* 60556 60655: gap of 100 bp
* 60656 62781: contig of 2126 bp in length
* 62782 62881: gap of 100 bp
* 62882 64370: contig of 1489 bp in length
* 64371 64470: gap of 100 bp
* 64471 65603: contig of 1133 bp in length
* 65604 65703: gap of 100 bp
* 65704 67667: contig of 1964 bp in length
* 67668 67767: gap of 100 bp
* 67768 69279: contig of 1512 bp in length

TITLE

JOURNAL

COMMENT

* 69280 69379: gap of 100 bp
* 69380 71150: contig of 1771 bp in length
* 71151 71250: gap of 100 bp
* 71251 73611: contig of 2361 bp in length
* 73612 73711: gap of 100 bp
* 73712 75366: contig of 1655 bp in length
* 75367 75466: gap of 100 bp
* 75467 78278: contig of 2813 bp in length
* 78280 78378: gap of 100 bp
* 78380 80372: contig of 1993 bp in length
* 80373 80472: gap of 100 bp
* 80473 82236: contig of 1764 bp in length
* 82237 82336: gap of 100 bp
* 82337 84113: contig of 1777 bp in length
* 84114 84213: gap of 100 bp
* 84214 86311: contig of 2102 bp in length
* 86316 86415: gap of 100 bp
* 86416 89969: contig of 3554 bp in length
* 89970 90069: gap of 100 bp
* 90070 92420: contig of 2351 bp in length
* 92421 92520: gap of 100 bp
* 92521 95426: contig of 2906 bp in length
* 95427 95526: gap of 100 bp
* 95527 97256: contig of 1730 bp in length
* 97257 97356: gap of 100 bp
* 97357 100888: contig of 3532 bp in length
* 100889 100988: gap of 100 bp
* 100989 103643: contig of 2655 bp in length
* 103644 103743: gap of 100 bp
* 103744 106545: contig of 2802 bp in length
* 106546 106645: gap of 100 bp
* 106646 111055: contig of 4410 bp in length
* 111056 111155: gap of 100 bp

Query Match	51.28;	Score 1719.8;	DB 2;	Length 149830;
Best Local Similarity	94.09;	Pred. No. 0;		
Matches 1748;	Conservative 0;	Mismatches 107;	Indels 4;	Gaps
Qy 931	ggggtccttcctccctgggtgctcaggaacccacagccacaagcccccctgcgaaggctcaggc	990		
Db 51471	GGGCTCCTTCCCTTGGTGCTCAGGAACCCACAGCCACAAGCCCCCTGCCAAGGCTCAGGC	51412		
Qy 991	agcctggccctgggaggactccagctctgttaggggcccctaaatgtctctcccacactg	1050		
Db 51411	AGCCTGGCCCCCTGGGAGACTCCACAGCTGTGTTAGGGGCCCTAAATGTCTCCCCACACTG	51352		
Qy 1051	tggtgtcccttctcttagtgtacacccctgtgtgctgtgggcatctgtgatggcag	1110		
Db 51351	TGGGTGCGCTTCTCTCTTAGTGTGCACCTGTGTGGCTGTGGGCATCTGTGCATGGCAG	51292		
Qy 1111	gccggggcgggcgatgctctgctgtctgtgatgggtatggagccgtctgttcatta	1170		
Db 51291	GCCGGGGCGGGGCATGTCTGGTGCTTGTGCTGATGGGTATGGACCGTCTGTTCATTA	51232		
Qy 1171	tgaagtgggtcagagctgtgattctgttgagcatgtgtgcatgcatgtgacctcat	1230		
Db 51231	TGAAGTGGGCTCAGAGCTGTGATTCTGTGAGCATGTGTGCATGTCATGTCAGCTCAT	51172		
Qy 1231	tgtccagtggtgaaagtgacatttccaaatctgagcattggacatcagtgatgtctctgtg	1290		
Db 51171	TGTCAGGTGGTGAAGGTGACATTTCCAAATCTGAGCATTTGGACATCAGTGTCTCTGTG	51112		
Qy 1291	tccctgtgtccctaccatccctgatggctgcaggagccgctgggcccctgccctcagtc	1350		
Db 51111	TCCCTGTGTCTCACCATCCCTGTATGCTGCAGGAGCCCGCTGGGCCCTTGCCCTCTAGTC	51052		
Qy 1351	acattccqccacctctggcacaggttggtgttctgcagtgaagacaaagcagagctgtg	1410		
Db 51051	ACATTCCCGCACCTCTGSCACAGGTTGGTGTTCTGTGCAGATGAAGACAAAGCGAGGCTG	50992		
Qy 1411	gattccagcatcttctctgaagccctggacagtcctgtgacgagacgaaagacctgagcc	1470		
Db 50991	GATCCAGCATCTTCTCTCAGCCCTTGACAGTCCTTGACAGTCCTTGACAGACGAAGACCTGAGCC	50932		

QY 2551 ggaaccacccctggcagcctgtggtatcggtttaaacacacagggaagcctgcagccccc 2610
Db 49855 GGGAAACACCTTGGCCAGCCTGTGGCTATCGTTTAAACACTGGGAAGGCTTCAGCCCCC 49796

QY 2611 agccgacagcttagctgcagctgacagcttccagctccagctagacagcttccaccc 2670
Db 49795 AGGCCGACAGTTAGTGTGACAGTGTGACAGTTCACAGTGCAGGTAGACGATGCTCCACCC 49736

QY 2671 acctcatgctgatctctctgtcatagtgaagtgtctggaacacaccttcacgttatg 2730
Db 49735 ACCTACTCANN 49676

QY 2731 ggaatctgtgtcccccagagtggtgcaatgaatggagtggaacagctcaactcgggtg 2789
Db 49675 NNN 49617

RESULT 14
HSNCF154
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SEGMENT
SOURCE
ORGANISM

HSNCF154 918 bp DNA PRI 07-JAN-1998
Homo sapiens p47-phox (NCF1) pseudogene, clone P38, exons 6 and 7.
U69642
U69642.1 GI:2754733

4 of 5
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B.,
Green,E.D., Chanock,S.J. and Curnutte,J.T.
A p47-phox pseudogene carries the most common mutation causing
p47-phox deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
97474758

JOURNAL
MEDLINE
REFERENCE
AUTHORS

Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T.,
Christensen,B., Curnutte,J.T. and Gorlach,A.
Characterization of the genomic structure of the p47-phox gene
Unpublished
3 (bases 1 to 918)
Gorlach,A., Lee,P.L., Roesler,J., Christensen,B., Chanock,S.J. and
Curnutte,J.T.
Direct Submission
Submitted (06-SEP-1996) Molecular and Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL 1, La
Jolla, CA 92037, USA
Location/Qualifiers
1. 918
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23"
/clone="P38"
/note="pseudogene is defined by the presence of a GT
deletion at the beginning of exon 2"
20. 143
/gene="NCF1"
/number=6
/pseudo
606. 713
/gene="NCF1"
/number=7
/pseudo

exon

BASE COUNT 163 a 261 c 293 g 201 t

ORIGIN

Query Match 26.4%; Score 886.4; DB 9; Length 918;
Best Local Similarity 99.1%; Pred. No. 8e-217; Gaps 2;
Matches 912; Conservative 0; Mismatches 6; Indels 2;

QY 768 tgccctccctctgtccccagacatcacccggccccaatcatctctgcagcgtaccgcgcaat 827
Db 1 TGCCCTCCCTCTTGTGCCCAGACATACCGGCCCATCATCTCTGCAGAGTACCGGCCAT 60

QY 828 tgccgaactacgagaagacctcgggtccagatgctctctgtccacggggaagctggtgga 887
Db 61 TGCCAACTACGAGAAGACCTCGGGCTCCGAGATGCTGTGTCCACGGGGGAGCGTGGTGA 120

QY 888 ggtcgtgagaagacgagcgggtcagacctccacaccttacgggggtctctccctcgtgt 947
Db 121 GGTCTGTGAGAAGACGAGAGCGGTTCAGACCTCCACCTTACGGGGCTCTTCCCTCGT 180

QY 948 gctcaggaaacacacagccacaaagccccctgccaaaggctcagcagcctggccccctggag 1007
Db 181 GCTCAGGAACCCACAGCCACAAAGCCCCCTGGCAA -GCTCAGGCAGCCTGGCCCCCTGGGAG 239

QY 1008 gactccagctctgttagggggccctaaatgtctctccacacctgtgggtcgccttctctct 1067
Db 240 GACTCAGGCTCTGTAGGGGCCCTAAATGTCTCTCCCACTGTGGGTGCGCTTCTCTCT 299

QY 1068 tagtgtcacccctgtgtggtgtgtggtcatctgtcatggcagccggggcggtgcatgt 1127
Db 300 TAGTGTGCACCCTGTGTGGCTGTGGGCTGTGTGCATGGCAGCGCGGGGGGCGCATGT 359

QY 1128 ctgctgttctgtctgggtgggtatgggacctctgttctaatatgaagtgggctcagagc 1187
Db 360 CTGCGTGTCTCTGTGGATGGGTATGGGACCTCTGTCTTCAATTATGAAGTGGCTCAGAGC 419

QY 1188 tgtgattctgtgagcatgtgtcatgcatgcatgtgacctcaattgtccagttggtgaag 1247
Db 420 TGTGATTCTGTGAGCATGTGTGCATGCATGCATGCATGCATGCATGCATGCATGCATGCAT 479

QY 1248 gtgacatttccaaatctgagcatgtgacatcagtggtctgtctgtccctgtgtcccaacca 1307
Db 480 GTGACATTTCCAAATCTGAGCAATGGACATTCGATGTGTGTGTGTGTGTGTGTGTGTGT 539

QY 1308 tccctgatggtgtcaggagccgctgggacctctgccccctcagtcacattcccgcacctgtg 1367
Db 540 TCCCTGATGGCTGCAGGAGCGCTGGGGCCCTGGCCCTCAGTCAGTCACATTCGCCGACCTCTG 599

QY 1368 gcaaggttgggtgtctgtcagatgaaagcaagcagagctggatccacagcatccttcc 1427
Db 600 GCACAGGTGTGGGT 659

QY 1428 tcgagccctgacagctcctgacgagcgaagaccctgagcccaactatgcaggtgccc 1487
Db 660 TCGACCCCTGGACAGTCTCTGACGACAGCAAGACCCCTGAGCCCAACTATGACAGGTGCC 719

QY 1488 cctgcctccgaggtgtgtaggggtgtgggagaaaggggcaggcaggtcaggggataattg 1547
Db 720 CCTGCCCTCCGAGGCTGTAGGGGTGTGGGAGAAAGGGGCAGGCAGGGCTCAGGGATATTG 779

QY 1548 agtgactccttggagctcgtgggt 1607
Db 780 AGTGACTGCTTGGAGTCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 839

QY 1608 tcctgcctcctgctgtgggtggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1667
Db 840 TCATCTCTCTGTGGCTTGGGGCCCTGGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 898

QY 1668 caggaggaggagcagacgac 1687
Db 899 CAGGAGGAGGAGCAGACGAC 918

RESULT 15
HS39NCF5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SEGMENT

HS39NCF5 904 bp DNA PRI 07-JAN-1998
Homo sapiens p47-phox (NCF1) pseudogene, clone P39, exons 6 and 7.
U72360
U72360.1 GI:2754740

5 of 6

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 904)
AUTHORS Gorlach,A., Lee,P.L., Roesler,J., Hopkins,P.J., Christensen,B., Green,E.D., Chanock,S.J. and Curnutte,J.T.
TITLE A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease
J. Clin. Invest. 100 (8), 1907-1918 (1997)
JOURNAL 97474758
MEDLINE 2 (bases 1 to 904)
REFERENCE Chanock,S.J., Roesler,J., Hopkins,P.J., Lee,P.L., Bassett,D.T., Christensen,B., Curnutte,J.T. and Gorlach,A.
AUTHORS Characterization of the genomic structure of the p47-phox gene Unpublished
3 (bases 1 to 904)
JOURNAL Gorlach,A., Lee,P.L., Chanock,S.J. and Curnutte,J.T.
AUTHORS Direct Submission
TITLE Submitted (24-SEP-1996) Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., CAL1, La Jolla, CA 92037
JOURNAL
FEATURES Location/Qualifiers
source 1..904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q11.23"
/clone="P39"
/note="pseudogene is defined by the presence of a GT deletion at the beginning of exon 2"
1..123
/gene="NCF1"
/number=6
/pseudo
586..893
/gene="NCF1"
/number=7
/pseudo
BASE COUNT 164 a 251 c 293 g 196 t
ORIGIN
Query Match 25.9%; Score 870.8; DB 9; Length 904;
Best Local Similarity 99.0%; Pred. No. 8.1e-213;
Matches 897; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
Qy 788 acataccggcccatcactctgcagacgtaccgcgcattgcgcgactacacgagacact 847
Db 1 ACATCAGCGGCCCATCATCTCTGCAGAGCTACCGCGCCATTGCCAACTACGAGAGACCT 60
Qy 848 cgggctcagagatgctctgtccacggggagcgtagtgtagggtcgtggagaagacgaga 907
Db 61 CGGGCTCCGAGATGGCTCTGTCCACGGGGGACGTGGTGGAGGTCGTGGAGAAGACGAGA 120
Qy 908 gcggctcagacctccacaccttaagggtcctctccctggtgctcaggaaacccacagccac 967
Db 121 CGGGTCAGACCTCCACCTTACGGGGCTCTCTCCCTCGTCTCAGGAACCCACAGCCAC 180
Qy 968 aagccccctgcagggtcaggcagcctggccccctggaggactccagctctgttagggg 1027
Db 181 AAGCCCCCTGCCAA-GCTCAGGACAGCTGGCCCCCTGGGAGGACTCAGGCTCTGTTAGGGG 239
Qy 1028 ccctaatagtctcccaactgtgggtgcgcttctctcttagtgtgcacctgtggtgg 1087
Db 240 CCGTAAGTGTCTCCACACTGTGGGTGGCTTCTCTTAGTGTGCACCTCTGTTGGTGG 299
Qy 1088 ctgtgggcatctgtgcagggcggcgggcgagcgtctcgtgtctgtctgtggtg 1147
Db 300 CTGTGGGATCTGTGCATGGGAGGCGGGGGGGGCATGTCTGCGGTGTCTGTCTGGATG 359
Qy 1148 ggtatgggacgctgttctattgaagtgggctcagagctgtgattctgtgagcatgtg 1207
|||||

Db 360 GGATGGGACCGTCTGTTCATTATGAAGTGGGCTCAGAGCTGTGATTCTGTGAGCATGTG 419
Qy 1208 tgcagcatgcatgtgacctcatgtgtccagtgagtgagtgagtgacatttccaaatctgaag 1267
Db 420 TGCATGCAATGATGACCTCATTTGCCAGTGTGTGTCCTCACCATCCCTGATGGCTGCAGGAG 479
Qy 1268 cattggacatcagtgctgtgtcctcctgtgtcctccacatccctcctgagtgagtgagggag 1327
Db 480 CATTGGACATCAGTGTGTCTGTGACCCCTGTGTCTCACCATCCCTGATGGCTGCAGGAG 539
Qy 1328 ccgctgggccccctgcctcagtcacattcccgcacacctctggcacagggttggtctctgt 1387
Db 540 CCGTGGGCCCTGCCCTCAGTCAATTCGCGACCTCTGGCACAGGTGGTGGTCTGTGT 599
Qy 1388 cagatgaagcaaacgaggtctggatcccagcatccttctcagacccctcagacagtcct 1447
Db 600 CAGATGAAGCAAAAGCGAGGCTGGATCCCGCATCTCTCTCGAGCCCCCTGGACAGTCTCT 659
Qy 1448 gacgagacggaagaccctgagcccccaactatgcaggtgccccctcctccagggtgtgtag 1507
Db 660 GACGAGACGGAAGACCCCTGAGCCCAACTATGCAGGTGCCCTGCCCTCCGAGGCTGTAG 719
Qy 1508 ggggtgggagaaagggcaggggctcaggggctcaggggatatgtgagtgactgttggaggtctg 1567
Db 720 GGGTGTGGGAGAAAGGGGCGCAGGCGCTCAGGGGATATTGAGTGACTGCTTGTGGAGTCTG 779
Qy 1568 ggcctggttgcctggccttgccagaaaagtcagggctaaagatcctcatcgctctggttgggg 1627
Db 780 GGTGTGTTGCTGGCTTGGCAGAAAAGTCAAGGCTAAGATCTCATCTCTCTGCGCTTGGGG 839
Qy 1628 gccctggcaggttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgc 1687
Db 840 GCCCTGGCAGGTTGTGTGTCCTTGGTCTGGACAGGAA-CAGGAGGAGGAGGAGCAGCAGAC 898
Qy 1688 tcggga 1693
Db 899 TCGGGA 904

Search completed: January 11, 2002, 17:18:54
Job time: 29486 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 11, 2002, 11:35:07.; Search time 8779.43 Seconds
(without alignments)
4113.771 Million cell updates/sec

Title: US-09-820-005-3_COPY_10200_13560
Perfect score: 3361
Sequence: 1 aaaggtcaggcactgtggc.....ctgtccaagtgggtgcatt 3361

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673.2	20.0	774	11	BG685989
2	362	10.8	15970	13	AQ839852
3	352.2	10.6	459	11	BF901739
4	350.2	10.4	743	10	AU130725
5	348.8	10.4	828	10	AL527073
6	348.4	10.4	732	13	AQ890095
7	347.4	10.3	1501	12	BC007465
8	345	10.3	658	13	AQ393450
9	342.2	10.2	680	13	AQ356404
10	340.4	10.1	372	11	BF900703
11	336.8	10.0	948	11	BG260565
12	335.6	10.0	877	13	AQ739838

13	332.2	9.9	947	11	BG680848
14	327.6	9.7	746	11	BF974962
15	327.4	9.7	710	13	AG010073
16	326	9.7	677	11	BG032943
17	324.6	9.7	778	10	AV764490
18	324.4	9.7	877	13	AQ739838
19	323.2	9.6	634	13	AQ543763
20	323.2	9.6	729	10	AV762129
21	322.6	9.6	705	10	AV760497
22	322.4	9.6	610	13	AQ479821
23	322.4	9.6	861	11	BG386803
24	321.8	9.6	679	11	BF346320
25	321.2	9.6	631	11	BE796439
26	320	9.5	1345	10	AV762220
27	318.6	9.5	728	10	AV763952
28	318.2	9.5	1345	10	AV762220
29	317.4	9.4	735	10	AL602867
30	316.8	9.4	851	10	AU122155
31	316.6	9.4	800	10	AU120942
32	316.4	9.4	922	10	AL515875
33	315.8	9.4	858	10	AU119532
34	315.6	9.4	948	11	BG260565
35	314.6	9.4	780	10	BE538259
36	314.4	9.4	6146	13	AQ839831
37	312.8	9.3	822	13	AQ748222
38	312.6	9.3	740	13	AQ878530
39	312.4	9.3	658	13	AQ393450
40	312.4	9.3	839	10	AV700988
41	312	9.3	1875	12	BC009270
42	311.8	9.3	759	10	AW976010
43	311.4	9.3	711	10	AV714931
44	310.4	9.2	891	11	BG744641
45	310	9.2	711	10	AL044340

ALIGNMENTS

RESULT 1
LOCUS BG685989 774 bp mRNA EST 01-MAY-2001
DEFINITION 602638552F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766245 5', mRNA sequence.
ACCESSION BG685989
VERSION BG685989.1 GI:13917386
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 774)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: Lfcm1626 row: g column: 14
High quality sequence stop: 773.
Location/Qualifiers
1. 774

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766245"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"


```
/db_xref="taxon:9606"
/clone_lib="MT0190"
/dev_stage="adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      77 a 125 c 109 g  61 t
ORIGIN
```

```
Query Match      10.1%; Score 340.4; DB 11; Length 372;
Best Local Similarity 99.4%; Pred. No. 1.2e-28;
Matches 352; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 829 gccgactcagagaacacccctcggctccgagatggtctgtccacgggggacgtggtgag 888
Db 372 gccgactcagagaacacccctcggctccgagatggtctgtccacgggggacgtggtgag 313

Qy 889 gtcgtgggaagagcagagcgttcagaccctccacacctcaggggtcctccctccctggtg 948
Db 312 gtcgtgggaagagcagagcgttcagaccctccacacctcaggggtcctccctccctggtg 253

Qy 949 ctcaggaacccacacacacacacacacacacacacacacacacacacacacacacacac 1008
Db 252 ctcaggaacccacacacacacacacacacacacacacacacacacacacacacacacac 193

Qy 1009 actccagctctgttaggggcccctaaatgctctccacacactgtgggtcgtctctctctt 1068
Db 192 actccagctctgttaggggcccctaaatgctctccacacactgtgggtcgtctctctctt 133

Qy 1069 agtgcacacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1128
Db 132 agtgcacacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 74

Qy 1129 tgcgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1182
Db 73 tgcgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 20
```

```
RESULT 11
LOCUS      BG260565      948 bp      mRNA      EST      13-FEB-2001
DEFINITION 602372119p1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4480109 5',
            mRNA sequence.
ACCESSION  BG260565
VERSION    BG260565.1 GI:12770381
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 948)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10313 row: e column: 06
            High quality sequence stop: 613.
            Location/Qualifiers
                1. 948
                    /organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_93"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      276 a 239 c 268 g 165 t
ORIGIN
```

```
Query Match      10.0%; Score 336.8; DB 11; Length 948;
Best Local Similarity 74.8%; Pred. No. 1.7e-28;
Matches 510; Conservative 0; Mismatches 122; Indels 50; Gaps 5;

Qy 1907 ttatttatttttttttttttttttttttttttttttttttttttttttttttttttttt 1966
Db 635 TTTGTTGTTTCTTTTGGAGATCAAGTTTGTCTTGTTCGCCAGGCTGGAGTGC 576

Qy 1967 ggtggtgtatctcagctcactgcaacctctgtctccacagg-atcaagtgtattctccagc 2025
Db 575 AATGGCGCAATCTCGGCTCAGTGCACACCTCCACCTCTGGGTTTCAAGCAATTTCTCTGC 516

Qy 2026 ctacgctccccagtagctgggtattacaagcatcacaccacacctggtgataattttg 2085
Db 515 CTCAGCCTC-CCAAGTAGCTGGGACTACAGGATGCGGCCACCATGCCAGCTAATTTTGG 457

Qy 2086 tatttttgttgagcgggttttaccagggttggccagactggtctcgaaactctgacct 2145
Db 456 TATTTTGTAGTAGACGGGGTTTACCATGTTGGCCAGGCTGCTCGAACCTCCGAGCT 397

Qy 2146 cagggtgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2205
Db 396 TAGGTGATCTGCGCCCTGGGCTCCCAAGTCTGGGATTACAGGAGTGAGCTACTGTG 337

Qy 2206 cctggtcctgggtttaccacacatttttaaaatggagtgtttcacccttttatgtgattaca 2265
Db 336 CCGGCCAG-----TGCCCTG 322

Qy 2266 gctgttttttttttttttttttttttttttttttttttttttttttttttttttttt 2325
Db 321 CTTTTTTTGTGTTGTTGGGATGGAGTCTAGCTCTGTCCACCTAGGCTGGAGTGCAG 262

Qy 2326 taatgcaactcagctcactgcaaccttagcctctgggttcaagcaaatctctcctgcctc 2385
Db 261 TGGCCCGATCTCGGCTCAGTGCACACTCCGCTCCAGGTTCAAGGATTTCTCTGCGCTC 202

Qy 2386 agccacctgtagctggtggtttacaggcatgacacacacacacacacacacacacacacac 2445
Db 201 AGCCCTCCTGAGT-GGCTGGGATTTCAAGGTGATCGCCACCATGCCAGCTAATTTTGTAT 143

Qy 2446 ttttttagtagagatgggttttcgcatgttggcagggtggtctcgaactcgtgacctca 2505
Db 142 TTTTGGTAGAGACAGAGATTTTTCACCATGTTGCTCAGGCTGGTCTCGAACTCCTGACCTCA 83

Qy 2506 ggtgatacgcgcgccttggtcctcccaaaagtgtgattacaggtgggaacacaccttgc 2565
Db 82 --TGATCTGCCACCTTGGCCTCCCAAAAGTGTGGGATTACAGGCTGAGCCACCTTGC 25

Qy 2566 cagcctgtggtatcgttttaa 2587
Db 24 CAGCCAGTGCCTGCTTTTGA 3
```

```
RESULT 12
LOCUS      AQ739838      877 bp      DNA      GSS      16-JUL-1999
DEFINITION HS_5505_Al_C09_47A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1081 Col=17 Row=E, DNA sequence.
ACCESSION  AQ739838
```


Db	342	-----TTACTGGTTGGGTGCGGT-CACGCTTTATAATCCACGACACTTTGGGAGGCC	394
Qy	359	gaggtgggagaatcctttgagccagaggttcgagaccagcctgggcatgtgcccaagac	418
Db	395	AGCTTAGCGGATCCCTGAGGTCAGGAGTTCAAGACAACGCTGGCCACATGCTGAAC	454
Qy	419	cttgtctttac-----aaaaattagccgggtgtggtggcatacgtctgtggtccca	470
*Db	455	CCTGCTCTACTATAAAATATAAAATTTAGCTGGCGGTGGCGAGGACCTGTAGTCCCA	514
Qy	471	gctattcgggaggtcaggcaggagattgcttgagcctaggaggtctagggtctgtagtga	530
Db	515	GCTAGTTGGGAGGCTGAGCAAGGGAATTCCTTTGAACCTGGGAGGCAAGGTTGCAGTGA	574
Qy	531	gctgtgacgtcactgactctagctcgtggcagacagacagacagactctgtcccaaaa	590
Db	575	GCTAGATTGCACCACGTACTCCAGCCTGGGCATCAAAAGTAGAGACTCTGCTTCAAAAA	634
Qy	591	agaaaaataaa 600	
Db	635	AACAAACAA 644	
RESULT	13		
LOCUS	BG680848	947 bp mRNA	EST
DEFINITION	602628648F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753762	5	01-MAY-2001
ACCSSION	BG680848	mRNA sequence.	
VERSION	BG680848.1	GI:13912245	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: James Cleaver, M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: L1AM10613 row: O column: 11		
	High quality sequence stop: 822.		
FEATURES	source		
	1. .947		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4753762"		
	/clone_lib="NCI_CGAP_Skn4"		
	/tissue_type="squamous cell carcinoma"		
	/lab_host="DH10B (TI phage-resistant)"		
	/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.		
	Average insert size 1.5kb. Library constructed by Life		
	Technologies. Note: this is a NCI_CGAP Library."		
BASE COUNT	276 a 203 c 260 g 208 t		
ORIGIN			
	Query Match 9.9%; Score 332.2; DB 11; Length 947;		
	Best Local Similarity 74.7%; Pred. No. 5.4e-28;		
	Matches 461; Conservative 0; Mismatches 143; Indels 13; Gaps 3		
Qy	5	ggtcaggcactgtgctcactgttaatcccgagcacttgggagcgagcggttgaa 64	
Db	43	GGCCAGATGTTGGGTCTATGCTCTATAATCCAGACACTTTGGGAGGCTGAGTAGGCAGA 102	
Qy	65	tcacctgaggtcaggaggttcgagaccagcctgaccaaatactgaaactccatctctact 124	
Db	103	TCACCTGAGGTGAGGAGTTCGAGACAACCTCGACAACATGGCAACACCTGCTCTACT 162	
Qy	125	aaaaatacaaaaattagccaggtgtgtgtggcgggtgtgtgttagtcccgactacttggag 184	
Db	163	AAAAATACAAAAATTAGCGGGGATGCTGGCAGCTGCCCGCTATTCGCCAGTACTCGGGAG 222	
Qy	185	gctgagcgagggtgaattacttgaacctggaggtggaggttgcaatgacgaagattgca 244	
Db	223	GCTGAGCGAGGAGATCGGTTTGAACTTGGGAGGAGGTTGCGGTGAGCCGAAGATCGTG 282	
Qy	245	ccactgcactcca-----gtacagagcgagactcctcatctcaaaaaaataaaaaa 298	
Db	283	CCACTGCACCTCCAGCCTGGGTACAGAGTGACACTCCATCTCAAAAAAATAAAAAA- 341	
Qy	299	agttggaaaaagccaggtgtcagttggtcctcagcctgttaatcccaacacactttaaggct 358	

[illegible]

Query Match 10.0%; Score 335.4; DB 3; Length 53526;
Best Local Similarity 75.7%; Pred. No. 1.9e-70;
Matches 514; Conservative 0; Mismatches 116; Indels 49;

Query Match	10.0%;	Score 335.4;	DB 3;	Length 53526;
Best Local Similarity	75.7%;	Pred. No. 1.9e-70;		
Matches 514;	Conservative	0;	Mismatches 116;	Indels 49;
				Gaps 6;

Db 7351 TGAGCCACGCGCCAGAC 7369

RESULT 6

US-08-814-095-7/c

; Sequence 7, Application US/08814095

; Patent No. 6025183

; GENERAL INFORMATION:

; APPLICANT: Soreq, Hermona

; APPLICANT: Zakut, Haim

; APPLICANT: Shani, Moshe

; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR

; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KOHN & ASSOCIATES

; STREET: 30500 No. 6025183Western Highway, Suite 410

; CITY: Farmington Hills

; STATE: Michigan

; COUNTRY: U.S.

; ZIP: 48334

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/814.095

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Montgomery, Ilene N.

; REGISTRATION NUMBER: 38,972

; REFERENCE/DOCKET NUMBER: 2391.00066

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (248) 539-5050

; TELEFAX: (248) 539-5055

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35060 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; DESCRIPTION: /desc = "Cosmid including ACHE

; DESCRIPTION: promotor, ACHE gene and ARS gene"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT: 7q22

; FEATURE:

; NAME/KEY: promotor

; LOCATION: 4089..22464

; OTHER INFORMATION: /function= "ACHE promotor"

; OTHER INFORMATION: /standard_name= "ACHE Promotor"

; FEATURE:

; NAME/KEY: exon

; LOCATION: 22465..22537

; OTHER INFORMATION: /function= "non-translated"

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 1

; FEATURE:

; NAME/KEY: exon

; LOCATION: 24090..25177

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /function= "(translation start:

; OTHER INFORMATION: 24110)"

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 2

; FEATURE:

; NAME/KEY: exon

; LOCATION: 25524..26009

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 3

; FEATURE:

; NAME/KEY: exon

; LOCATION: 27005..27274

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 4

; FEATURE:

; NAME/KEY: exon

; LOCATION: 27255..28007

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 5

; FEATURE:

; NAME/KEY: terminator

; LOCATION: 27385..27387

; FEATURE:

; NAME/KEY: exon

; LOCATION: 28008..28129

; IDENTIFICATION METHOD: experimental

; OTHER INFORMATION: /evidence= EXPERIMENTAL

; OTHER INFORMATION: /gene= "ACHE"

; OTHER INFORMATION: /number= 6

; FEATURE:

; NAME/KEY: terminator

; LOCATION: 28129..28131

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (34528..34895)

; OTHER INFORMATION: /function= "arsenite resistance

; OTHER INFORMATION: gene"

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 1

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (34092..34358)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 2

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33779..33963)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 3

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33493..33591)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 4

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (33297..33408)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 5

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32959..33094)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 6

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32569..32628)

; OTHER INFORMATION: /gene= "AR"

; OTHER INFORMATION: /number= 7

; FEATURE:

; NAME/KEY: exon

; LOCATION: complement (32386..32468)

GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, S.
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,963B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4421 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: JT101
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.1 kb Bam HI
OTHER INFORMATION: fragment derived from human placental
OTHER INFORMATION: genomic DNA
US-08-157-963B-9

Query/ Match 9.7%; Score 326.6; DB 2; Length 4421;
Best Local Similarity 75.4%; Pred. No. 8.9e-69;
Matches 500; Conservative 0; Mismatches 99; Indels 64; Gaps 5;

Qy 1012 tattatttacttttttgagatgagatctcactcccttgctgaaggtgagagtcggtg 1971
Db 3491 TATTATTTTCTTTTGGAGATGAGTCTGGCTCTGTCAACCCAGGCTGGAGTCAGTGG 3550

Qy 1072 tgcattctcagctcactgcaacctctgctccaggatcgaagtattctccagctcagc 2031
Db 3451 CGCGATCTCGGCTCAGTCAAGCTCTTCTCCCGGGTTCACCCCATCTCTCGCTCAAC 3610

Qy 2032 ctccagggtgagctgggattacaagcatgcaccaccatgcctggataa--tttttgtatt 2089
Db 3411 CTCCCGA-GTAGCTGTGAATACAGCGGTCCGCCACCCAGCGCTGATTTTGTATT 3669

Qy 2090 tttagtgagacgggggttttaccagggttgccagactgggtctgaacttttgacctcagg 2149
Db 3470 TTTAGTAGAGACGGGGTTTACCGGTGTAGCCAGGATGGTCTGGATCTCCTGACCTC--G 3727

Qy 2150 tgatctgctgctcctgcgctcccaaaagtgtggaattacagatgtgagccactgtccctg 2209
Db 3728 TGATCTGCTGCTCGGCTCCCAAAAGTGTGGGATACAGGCTTGAGCCACCGCACCG 3787

Qy 2210 gcctgggttaccacacattttaaaatggagtgatttcaaccttttatgtgattacagctt 2269
Db 3788 GCC----- 3790

Qy 2270 gtt 2329
Db 3791 -TCTATTTT----- 3849

Qy 2330 gcaatctcagctcactgcaacctctagcctcctgggttcaagcaattctcctgcctcagcc 2389
Db 3850 CGATCTCGGCTCACTGCAACCTCCGCTCTCTGGGTTCAAGAGATTCTCTGCTCAGCC 3909

Qy 2390 acctgagtagcctggggtttacagatgcaccacagccagggtaatttttttttttttttt 2449
Db 3910 TCCCAAGTAG-CTGGGATTACAGGTGCCACACACACCGCTGGCTAGTTTGTATTTT 3968

Qy 2450 tagtagagatggggttttcgcccattgtggcaggtgtgtctcgaactcctgacctcaggtg 2509
Db 3969 TAGTAAAGATGGGTTTACCATGTTGCCAGGCTGGTCTTGAACCTCCTGACATCAGGTG 4028

Qy 2510 atccgcccgccttgccctcccaaaagtctagattacaggtgggaaccaccttgcccagc 2569
Db 4029 ATCCGCCACCTTAGCTCCCAAAAGTCTGGGATTACAGGCGTGAGCCACCATACCTGGC 4088

Qy 2570 ctg 2572
Db 4089 CAG 4091

RESULT 10
PCT-US95-07201-9
Sequence 9, Application PC/TUS9507201
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, Sofia
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434

REFERENCE/DOCKET NUMBER: 20264126PCT

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-8849

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 4421 Base Pairs

TYPE: Nucleic Acid

STRANDEDNESS: Double

TOPOLOGY: Unknown

MOLECULE TYPE: Genomic DNA

ORGANISM: Human

FEATURE:

NAME/KEY: JT1

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION: 7.1 kb Bam HI

OTHER INFORMATION: fragment Derived from human placental

OTHER INFORMATION: genomic DNA: Also referred to as JT101

PCT-US95-07201-9

Query Match 9.7%; Score 326.6; DB 5; Length 4421;
Best Local Similarity 75.4%; Pred. No. 8.9e-69;
Matches 500; Conservative 0; Mismatches 99; Indels 64; Gaps 5;

```
Qy 1912 tattatttacttttttgagatgagtgctcactccctgtctaaggtcgagtgcggtg 1971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3491 TATTTTCTTTTCTTTGAGATGAGTCTGGCTCTGCACCCAGGCTGGAGTGG 3550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1972 tctatctcagctcactcaacctctgctccagatcaagtgattctccagctcagc 2031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3551 CCGGATCTCGGCTCACTGCAAGCTCTTCTCCGGGTTTCACCCAATTCCTGCTCAAC 3610
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2032 ctccccaggtagctgggattcaagcatgcccaccatgcctggataa--ttttgtatt 2089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3611 CTCCCGA-GTAGCTGTAATACAGCGCTCCGCCACCCAGCCGCGACTAATTTTGTATT 3669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2090 tttagttgagcggggtttaccagggttggccagactgggtctgaactctgacctcagg 2149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3670 TTTAGTAGAGAGGGGTTTACCGTGTAGCCAGATGGTCTGGATCTCCCTGACCTC--G 3727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2150 tgatctgctgcctgcctcccaagtgctggaattacagatgtgagccactgcccctg 2209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3728 TGATCTGCTGCTCGGCTCCCAAAGTCTGGGATTTACAGGCTTGACCCACCGACCCG 3787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2210 gctgggttaccacacattttaaaatggagtgatttcaccccttttatgtggtattacagctt 2269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3788 GCC----- 3790
Qy 2270 gtttttttttttttttgagacaagctggctctgtcaaccaggctgagtgcaagtaat 2329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3791 -TCATTATTTTCTTTTGTAGATGAGTCTCACTGTCACTGTCACTGTGGCTGGAGTGGAGTGG 3849
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2330 gcaatctcagctcactgcaaccttagctctctgggttcaagcaattctcctgcctcagcc 2389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3850 CCGATCTCGGCTCACTGCAACCTCCGCCCTCTGGGTTCAAGAGATTCCTCTGCCCTCAGCC 3909
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2390 acctgagtagctggggtttacagggcatgcaccacacacagccagggttaattttgtatttt 2449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3910 TCCCAAGTAG-CTGGGATTACAGGTGCCACACACACCGCTGGCTAGTATTTTGTATTTT 3968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2450 tagtagagatggggttttcgcaatgttgccaggctgggtctcgaactcctgacctcagtg 2509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3969 TAGTAAGATGGGTTTCACTGTTGGCCAGGCTGGTCTTGAACCTCTGACATCAGGTG 4028
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2510 atccgccgccttggtctcccaagtgtagattacaggtgggaacacaccttggccac 2569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4029 ATCCGCCACCTTACCTCCCAAAGTGTGGGATTTACAGGCGTGAGCCACCATACCTGGC 4088
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2570 ctg 2572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 4089 CAG 4091

RESULT 11

US-09-009-913-1

Sequence 1, Application US/09009913

Patent No. 6087485

GENERAL INFORMATION:

APPLICANT: Axys Pharmaceuticals, Inc.

TITLE OF INVENTION: Asthma Related Genes

NUMBER OF SEQUENCES: 339

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bozicevic & Reed, LLP

STREET: 285 Hamilton Ave, Suite 200

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/009,913

FILING DATE: 21-JAN-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J

REGISTRATION NUMBER: 36,677

REFERENCE/DOCKET NUMBER: SEQ-4P

TELEPHONE: 650-327-3231

TELEFAX: 650-327-3231

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 72928 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-009-913-1

Query Match 9.6%; Score 321.8; DB 3; Length 72928;
Best Local Similarity 73.9%; Pred. No. 3.9e-67;
Matches 500; Conservative 0; Mismatches 122; Indels 55; Gaps 5;

```
Qy 1924 ttttttgagatagatctcactcccttgctaaggtgagtgagtggtgtctatctcagc 1983
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47733 TTTTCATGACAAAGTCTTGCTCTGTTGCCAGGCTGAGTGCAGTGGCGTATCTCTGC 47792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1984 tcactgaaacctctgctccagagatcaagtgatctctccagctcagctccccaggtag 2043
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47793 TCACCTCAACCTCTGCCCTCCAGGTTCAAGTGACTGTCTCAGGCTCTGGGA-GTAG 47851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2044 ctgggattacaagcatgcacacacatgcctgataattttgtatttttagttgagacgg 2103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47852 CTGGGACTTACAGGTGCGTGCCACCAACCACTGGCTAATTTTGTATTTTGTAGAGATGG 47911
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2104 gttttcacaggttgccagactgctcgaacttctgacctcaggtgatctgcctgcct 2163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47912 GGTTTTGCATGTTGCCAGGCTGGTCTCGAATCTCTGGCCCTCAAGTATTTGACTGCCT 47971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2164 cggctcccaagtgtcgtgaattacagatgtgagccactgtccctggcctggtgtaccac 2223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47972 TGGCTCCCAAAGTCTGGAATTTACAGGCGTGAGCCACCATGCTCTGGACTGCTGA----- 48026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2224 atttttaaatggagtgatttcaccccttttatgtggatttacagctgtttttttttttt 2283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Search completed: January 11, 2002, 12:08:49
Job time: 10881 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:07:38 ; Search time 25.69 Seconds
(without alignments)
2197.786 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDTFIRHIALLGKRRFVP.....ADILNRCSESTKRKLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL17:**

- 1: sp-archaea:**
- 2: sp-bacteria:**
- 3: sp-fungi:**
- 4: sp-human:**
- 5: sp-invertebrate:**
- 6: sp-mammal:**
- 7: sp-mhc:**
- 8: sp-organelle:**
- 9: sp-phase:**
- 10: sp-plant:**
- 11: sp-rodent:**
- 12: sp-virus:**
- 13: sp-vertebrate:**
- 14: sp-unclassified:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2023	99.4	390	4 Q9BU90	Q9BU90 homo sapien
2	2015	99.0	390	4 O43842	O43842 homo sapien
3	2005.5	98.6	389	4 Q9UDV9	Q9UDV9 homo sapien
4	1717.5	84.4	391	6 Q9N0E8	Q9N0E8 turslops tr
5	1697	83.4	390	11 Q9J134	Q9J134 mus musculus
6	1689.5	83.0	389	11 Q9NM65	Q9NM65 rattus norv
7	1689	83.0	390	11 Q70144	Q70144 mus musculus
8	1411	69.3	276	4 Q9BX18	Q9BX18 homo sapien
9	1346.5	66.2	310	11 Q9JK56	Q9JK56 rattus norv
10	1290	63.4	254	4 Q9BX17	Q9BX17 homo sapien
11	504.5	24.8	1124	11 Q89032	Q89032 mus musculus
12	334.5	16.4	1031	4 Q9H462	Q9H462 homo sapien
13	275	13.5	940	4 Q43302	Q43302 homo sapien
14	197	9.7	339	6 Q9GM25	Q9GM25 turslops tr
15	188	9.2	1248	4 Q9NZM2	Q9NZM2 homo sapien
16	188	9.2	1676	4 Q9ULG4	Q9ULG4 homo sapien
17	188	9.2	1681	4 Q9NYG0	Q9NYG0 homo sapien
18	188	9.2	1696	4 Q9NZM3	Q9NZM3 homo sapien
19	185.5	9.1	550	4 Q9P2Q1	Q9P2Q1 homo sapien

20	183	9.0	464	4	O95062	O95062 homo sapien
21	179.5	8.8	1270	13	O42287	O42287 xenopus lae
22	178	8.7	348	4	Q9BU98	Q9BU98 homo sapien
23	177.5	8.7	1197	11	Q9Z0R5	Q9Z0R5 mus musculus
24	177.5	8.7	1658	11	Q9Z0R6	Q9Z0R6 mus musculus
25	176.5	8.7	248	4	Q15812	Q15812 homo sapien
26	172	8.5	1714	11	Q9Z0R4	Q9Z0R4 mus musculus
27	169.5	8.3	102	4	Q9NTM6	Q9NTM6 homo sapien
28	168.5	8.3	248	11	Q9R143	Q9R143 mus musculus
29	168	8.3	637	11	Q9JLQ0	Q9JLQ0 mus musculus
30	166.5	8.2	1217	11	Q9WVE9	Q9WVE9 rattus norv
31	166	8.2	1220	4	Q9UKN1	Q9UKN1 homo sapien
32	166	8.2	1220	4	Q9UET5	Q9UET5 homo sapien
33	166	8.2	1721	4	O95216	O95216 homo sapien
34	166	8.2	1721	4	O9UNK2	O9UNK2 homo sapien
35	164.5	8.1	639	4	Q9Y5K6	Q9Y5K6 homo sapien
36	164	8.1	239	11	Q9D747	Q9D747 mus musculus
37	164	8.1	1011	5	O61639	O61639 drosophila
38	161	7.9	641	11	O88903	O88903 mus musculus
39	161	7.9	1094	5	O61618	O61618 drosophila
40	160.5	7.9	1146	11	Q9WVE1	Q9WVE1 rattus norv
41	159.5	7.8	1097	5	Q9VIF7	Q9VIF7 drosophila
42	151	7.4	187	4	Q9UK60	Q9UK60 homo sapien
43	150.5	7.4	665	4	Q9NYR0	Q9NYR0 homo sapien
44	150	7.4	585	11	O9JKQ1	O9JKQ1 rattus norv
45	147.5	7.2	290	3	P78815	P78815 schizosacch

ALIGNMENTS

RESULT 1
Q9BU90
ID Q9BU90 PRELIMINARY: PRT: 390 AA.
AC Q9BU90;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE NEUTROPHIL CYTOSOLIC FACTOR 1 (47KD, CHRONIC GRANULOMATOUS DISEASE, AUTOSOMAL 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LYMPHOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: BC002816; AAH02816.1; ...
SQ SEQUENCE 390 AA; 44681 MW; 3D91EDC99A1B6417 CRC64;

Query Match	99.4%	Score 2023;	DB 4;	Length 390;
Best Local Similarity	99.0%;	Pred No. 4.2e-153;		
Matches 386;	Conservative 0;	Mismatches 0;	Indels 4;	Gaps 1;
Qy 1	MGDTFIRHIALLGKRRFVPSQHYVYMLVKWQDLSEKVVYRRFTEIYEFHKLKEMFPI 60			
Db 1	MGDTFIRHIALLGKRRFVPSQHYVYMLVKWQDLSEKVVYRRFTEIYEFHKLKEMFPI 60			
Qy 61	ENGAINPENRIIPHLPAKPFQDQRAAENRQCTLTLEYCSTLMSLPTKISRCPLHLDFFKV 120			
Db 61	ENGAINPENRIIPHLPAKPFQDQRAAENRQCTLTLEYCSTLMSLPTKISRCPLHLDFFKV 120			
Qy 121	RPDDLKLPDNTQTKPEYLMKPKCKSTATDITGPILOTYRAIANYEKTSSEWALSTG 180			
Db 121	RPDDLKLPDNTQTKPEYLMKPKCKSTATDITGPILOTYRAIANYEKTSSEWALSTG 180			
Qy 181	DVVEVVEKSESGWFCQMKAKRGWIPASPLELDSPEDEDPNPYAGFPYVAIKAYTAV 240			
Db 181	DVVEVVEKSESGWFCQMKAKRGWIPASPLELDSPEDEDPNPYAGFPYVAIKAYTAV 240			

```
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQSGQDVSAQAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGWVIRKDDVTGYFPSPMYLQSGQDVSAQAQRIKRGAPP 300
QY 297 RRSSTRNHSHIORSKRKLSQDAYRNRNVSFRFLQQRROARPGPQSPGSPLEEROTQSK 356
DB 301 RRSSTRNHSHIORSKRKLSQDAYRNRNVSFRFLQQRROARPGPQSPGSPLEEROTQSK 360
QY 357 POPAVPPRPSADLIILNRCSESTKRKLASAV 386
DB 361 POPAVPPRPSADLIILNRCSESTKRKLASAV 390

RESULT 2
O43842
ID O43842 PRELIMINARY; PRT; 390 AA.
AC O43842;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE P47-PHOX.
GN NCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97474758; PubMed=9329953;
RA Goerlach A., Lee P.L., Roesler J., Hopkins P.J., Christensen B.,
RA Green E.D., Chanock S.J., Curnutte J.T.;
RT "A p47-phox pseudogene carries the most common mutation causing p47-
RT phox-deficient chronic granulomatous disease.";
RL J. Clin. Invest. 100:1907-1918(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chanock S.J., Roesler J., Zhan S., Hopkins P., Lee P., Barrett D.,
RA Christensen B.L., Curnutte J.T., Goerlach A.;
RT "Genomic structure of the human p47-phox gene.";
RL Blood Cells Mol. Dis. 0:0-0(2000).
DR EMBL; U57835; AAB95193.1; -.
DR EMBL; U57833; AAB95193.1; JOINED.
DR EMBL; U57834; AAB95193.1; JOINED.
DR EMBL; AF184614; AAF34737.1; -.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS00002; SH3; 2.
SQ SEQUENCE 390 AA; 44695 MW; A19746E845B044C8 CRC64;

Query Match 99.0%; Score 2015; DB 4; Length 390;
Best Local Similarity 98.5%; Pred. No. 1.8e-152;
Matches 384; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 1 MGDFTIRHALLGFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHFKTKEMPEPI 60
DB 1 MGDFTIRHALLGFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHFKTKEMPEPI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFKVV 120
DB 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFKVV 120
QY 121 RPDDLKLPDNTQKKPETYLMKDKGKSTATDITGPTIIQLQYRAIANYEKTSGSEMASTG 180
DB 121 RPDDLKLPDNTQKKPETYLMKDKGKSTATDITGPTIIQLQYRAIANYEKTSGSEMASTG 180
QY 181 DVVEVSESGWFCQMKAKRGWIPASFLPDSDETEDEPNYAGEPYVAIKAYTAV 240
```

```
DB 181 DVVEVSESGWFCQMKAKRGWIPASFLPDSDETEDEPNYAGEPYVAIKAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW----KDDVTGYFPSPMYLQSGQDVSAQAQRIKRGAPP 296
DB 241 EGDEVSLLGEAVEVIHKLDDGWVIRKDDVTGYFPSPMYLQSGQDVSAQAQRIKRGAPP 300
QY 297 RRSSTRNHSHIORSKRKLSQDAYRNRNVSFRFLQQRROARPGPQSPGSPLEEROTQSK 356
DB 301 RRSSTRNHSHIORSKRKLSQDAYRNRNVSFRFLQQRROARPGPQSPGSPLEEROTQSK 360
QY 357 POPAVPPRPSADLIILNRCSESTKRKLASAV 386
DB 361 POPAVPPRPSADLIILNRCSESTKRKLASAV 390

RESULT 3
Q9UDV9
ID Q9UDV9 PRELIMINARY; PRT; 389 AA.
AC Q9UDV9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE WUCSC:H_DJ0771P04.4 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kalicki J., Lapidin Y.;
RT "The sequence of Homo sapiens PAC clone RP4-771P4.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004883; AAD15422.1; -.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR Pfam; PF00018; SH3; 2.
DR Pfam; PF00787; PX; 1.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 2.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
SQ SEQUENCE 389 AA; 44584 MW; 51BF3DA576497EEC CRC64;

Query Match 98.6%; Score 2005.5; DB 4; Length 389;
Best Local Similarity 98.7%; Pred. No. 1e-151;
Matches 385; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 1 MGDFTIRHALLGFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHFKTKEMPEPI 60
DB 1 MGDFTIRHALLGFEKRFVPSQHYVMFLVKWQDLSEKVVYRRFTEIYEFHFKTKEMPEPI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFKVV 120
DB 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCSLMSLPTKISRCPHLLDFKVV 120
```

```

QY 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
Db 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
QY 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240
Db 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRIKRGAPP 296
Db 241 EGDEVSLLGEAVEVIHKLDDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRIKRGAPP 296
QY 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEEROTQSK 356
Db 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEEROTQSK 356
QY 357 POPAVPPRPSADLILNRCSESTRKRLASAV 386
Db 360 POPAVPPRPSADLILNRCSESTRKRLASAV 389

RESULT 4
Q9N0E8 PRELIMINARY: PRT: 391 AA.
AC Q9N0E8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P47-PHOX.
GN Tursiops truncatus (Atlantic bottle-nosed dolphin).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Tursiops.
OX NCBI_TaxID=9739;
RN [1]
RP SEQUENCE FROM N.A.
RT Inoue Y., Ito T., Sakai T.;
RT "Molecular Cloning and Identification of Bottle-Nosed Dolphin
RT P47phox.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035594; BAA36544.1;
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
SQ SEQUENCE 391 AA; 44766 MW; F09E96D6B74B7C7C CRC64;

Query Match 84.4%; Score 1717.5; DB 6; Length 391;
Best Local Similarity 82.7%; Pred. No. 8.8e-129;
Matches 324; Conservative 30; Mismatches 31; Indels 7; Gaps 3;

QY 1 MGDFTIRHIALGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
Db 1 MGDFTIRHIALGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGATNPENRIIPLPAKWFQDQRAENROCTLTFCSTLMSLPTKISRCPHLLDFKVV 120
Db 61 EAGATNPENRIIPLPAKWFQDQRAENROCTLTFCSTLMSLPTKISRCPHLLDFKVV 120
QY 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
Db 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
QY 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240
Db 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240

```

```

Db 181 DVVDVVEKSSQSWFCOMKTKRGWVPASYLEPLDSPDEAEDPEPNYEGEPTIIEAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRIKRGAPP 296
Db 241 LEDEISLQGETIEVIHKLDDGWVIRKEDVTGYFPSPMYLQQAQGDQDAQRIKRGAPP 300
QY 297 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEER--QTOR 354
Db 301 RRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEER--QTOR 354
QY 355 SKPQAVPPRPSADLILNRCSESTRKRLASAV 386
Db 360 PKPQAVPPRPSADLILNRCSESTRKRLASAV 391

RESULT 5
Q9J134 PRELIMINARY: PRT: 390 AA.
AC Q9J134;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P47-PHOX.
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RT Green E.D.;
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF267747; AAF90134.1;
DR MGD; MGI:97283; Ncfl.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
SQ SEQUENCE 390 AA; 44667 MW; C8EFAB953839CE9A CRC64;

Query Match 83.4%; Score 1697; DB 11; Length 390;
Best Local Similarity 81.1%; Pred. No. 3.8e-127;
Matches 317; Conservative 34; Mismatches 34; Indels 6; Gaps 3;

QY 1 MGDFTIRHIALGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
Db 1 MGDFTIRHIALGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGATNPENRIIPLPAKWFQDQRAENROCTLTFCSTLMSLPTKISRCPHLLDFKVV 120
Db 61 EAGATNPENRIIPLPAKWFQDQRAENROCTLTFCSTLMSLPTKISRCPHLLDFKVV 120
QY 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
Db 121 RPDDLKLPDNTQKPEYLMFKDCKSTATDITGPIILQTYRAIANYEKTSSEMSALSTG 180
QY 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240
Db 181 DVVEVVESESGWFCOMKAKRGWIPASFLPLDSDPDETEDEPNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLGEAVEVIHKLDDGW---KDDVTGYFPSPMYLQKSGQDVSOAQRIK--RGAP 295
Db 241 EDEMSLSEGEAIEVIHKLDDGWVIRKEDVTGYFPSPMYLQKAGEITQAOQRIKRGAP 300
QY 296 PRSIRNAHSIHQRKRKLSQDAYRRNSVRFLOQRRRQARPGPQSPGSPLEEROTQSK 355

```


Db 301 PRSTIRNAQSIHQSRKRLSQDTYRNRSVRFLOQRRRGRGPLSTDG-TKDNFSTPRV 359
QY 356 KPQAVPPRPSADLIILNRCSESTKRLASAV 386
Db 360 KPQAVPPRPSADLIILHRCSTKRLKLSAV 390

RESULT 6
Q99M65 PRELIMINARY; PRT; 389 AA.
ID Q99M65;
AC Q99M65;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P47 PHOX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Tanabe M., Radmark O.P.;
RT "Full length sequence of rat p47 phox.";
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY029167; AAK31797.1; -.
SQ SEQUENCE 389 AA; 44742 MW; BFC75842E53E68A4 CRC64;

Query Match 83.0%; Score 1689.5; DB 11; Length 389;
Best Local Similarity 81.1%; Pred. No. 1.5e-126;
Matches 317; Conservative 34; Mismatches 33; Indels 7; Gaps 3;

QY 1 MGDFTIRHIALLGFEKRFVPSOHYVYMFVLVKKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
Db 1 MGDFTIRHIALLGFEKRFVPSOHYVYMFVLVKKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCTSLMSLPTKISRCPHLLDFFKV 120
Db 61 EAGAIHTENRVIPHPAPRWYDQRAAESRQGLTEYFNLSLMLPKISRCPHLLNFKV 120
QY 121 RPDDLKLPDNTQKPTVLMKDGKSTATDITGPILQTYRAIANYEKTSSEMASTG 180
Db 121 RPDDLKLPDNTQKPTVLMKDGKSTATDITGPILQTYRAIANYEKTSSEMASTG 180
QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
Db 181 DVVDVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLEGEAVEVTHKLLDGW---KDDVTGYFSPMYLQKSGDVSOAQRQIK-RGAP 295
Db 241 EEDEVSLSEGEALVTHKLLDGWVWVVRKGDITGYFSPMYLQKAGEITQAOQRIRSGAP 300
QY 296 PRSSTIRNAHSIHQSRKRLSQDAYRNRSVRFLOQRRRQARPGPSGSPLEERQOTRS 355
Db 301 PRSTIRNAQSIHQSRKRLSQDTYRNRSVRFLOQRRRQARPGPSGSPLEERQOTRS 358
QY 356 KPQAVPPRPSADLIILNRCSESTKRLASAV 386
Db 359 KPQAVPPRPSADLIILHRCSTKRLKLSAV 389

RESULT 7
O70144 PRELIMINARY; PRT; 390 AA.
ID O70144;
AC O70144;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P47PHOX.
GN NCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEUKEMIA;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
DR EMBL; AB002663; BAA35649.1; -.
DR MGD; MGI:97283; NCF1.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001883; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00498; P47PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS00002; SH3; 2.
SQ SEQUENCE 390 AA; 44698 MW; A383DB953839CFBC CRC64;

Query Match 83.0%; Score 1689; DB 11; Length 390;
Best Local Similarity 80.8%; Pred. No. 1.6e-126;
Matches 316; Conservative 34; Mismatches 35; Indels 6; Gaps 3;

QY 1 MGDFTIRHIALLGFEKRFVPSOHYVYMFVLVKKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
Db 1 MGDFTIRHIALLGFEKRFVPSOHYVYMFVLVKKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
QY 61 EAGAINPENRIIPHPAPKWFQDQRAAENRQGLTEYCTSLMSLPTKISRCPHLLDFFKV 120
Db 61 EAGAIHTENRVIPHPAPRWYDQRAAESRQGLTEYFNLSLMLPKISRCPHLLDFFKV 120
QY 121 RPDDLKLPDNTQKPTVLMKDGKSTATDITGPILQTYRAIANYEKTSSEMASTG 180
Db 121 RPDDLKLPDNTQKPTVLMKDGKSTATDITGPILQTYRAIANYEKTSSEMASTG 180
QY 181 DVVEVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
Db 181 DVVDVVEKSESGWFCOMKAKRGWIPASPLEDSDPEDEPNNYAGPYVAIKAYTAV 240
QY 241 EGDEVSLLEGEAVEVTHKLLDGW---KDDVTGYFSPMYLQKSGDVSOAQRQIK-RGAP 295
Db 241 EEDEVSLSEGEALVTHKLLDGWVWVVRKGDITGYFSPMYLQKAGEITQAOQRIRSGAP 300
QY 296 PRSSTIRNAHSIHQSRKRLSQDAYRNRSVRFLOQRRRQARPGPSGSPLEERQOTRS 355
Db 301 PRSTIRNAQSIHQSRKRLSQDTYRNRSVRFLOQRRRQARPGPSGSPLEERQOTRS 359
QY 356 KPQAVPPRPSADLIILNRCSESTKRLASAV 386
Db 360 KPQAVPPRPSADLIILHRCSTKRLKLSAV 390

RESULT 8
Q9BX18 PRELIMINARY; PRT; 276 AA.
ID Q9BX18;
AC Q9BX18;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ADAPTOR PROTEIN P47PHOX (FRAGMENT).
GN NCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

Matches	250;	Conservative	27;	Mismatches	28;	Indels	5;	Gaps
QY	21	SOHYVMFLVKWQDLSEKVVYRRFTETIYEFHKTLMEMFTEAGAINPENRIIPHLPAKWK	80					
Db	1	SOHYVMFLVKWQDLSEKVVYRRFTETIYEFHKTLMEMFTEAGAINPENRIIPHLPAKWK	60					
QY	81	FDQRAAENRQGLTTEYCYSTLMSLPTKISRCPHLLDFFKVRPDDLKLPDNTQTKKPEYTL	140					
Db	61	YDQRAAESRQGLTTEYFNSLMLPKISRCPHLLNFFKVRPDDLKLPDNDQVKKPEYTL	120					
QY	141	MPKDGKSTADLTGPIILQTYRAIANYEKTSGSEMALSTGDDVVVEKSESQWFWCQMA	200					
Db	121	TAKDGKNNVADINSGPIILQTYRAIADYEKSGKTEMTATGDDVVVEKSESQWFWCQMP	180					
QY	201	KRCWIPASFLPLELSDPEDTEDPNPVAGEPYVAIKAYTAVEGDEVSLLEGEAEVEIHKLL	260					
Db	181	NRGWVPASYLEPLSDPEADDPNPVAGEPYVTIKAYAAVEDEVSLSEGEATEVIHKLL	240					
QY	261	DGW-----KDDVTGYFSPMYLQKSGQDVVSQAQROIK-RGAPPRSSIRNAHSIHQSRKRL	315					
Db	241	DGWWVVRKGDITGYFSPMYLQKAGEKITQAQRIIRSRGAPPRSTIRNAQSIHQSRKRL	300					
QY	316	SQDAYRRNSV 325						
Db	301	SQPTYRRNSV 310						
RESULT	10							
Q9BXI7		PRELIMINARY; PRT; 254 AA.						
ID	Q9BXI7							
AC	Q9BXI7							
DC	01-JUN-2001 (TrEMBLrel. 17, Created)							
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)							
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)							
DE	ADAPTOR PROTEIN P47PHOX (FRAGMENT).							
GN	NCFl.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
RP	[1]							
RN	SEQUENCE FROM N.A.							
RC	TISSUE=UMBILICAL VEIN;							
RA	Gu Y., Xu Y., Souza R.F., Nwariaku F.E., Terada L.S.;							
RL	"Activation of c-Jun amino terminal kinase by a signalling oxidase.";							
RT	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AF30626; AAK19517.1; -.							
FT	NON_TER 1							
FT	VARIANT 134 134							
FT	NON_TER 254 254							
SQ	SEQUENCE 254 AA; 28736 MW; FD7CC225ED7EDCF3 CRC64;							
Query Match	63.4%;	Score 1290;	DB 4;	Length 254;				
Best Local Similarity	97.6%;	Pred. No. 5.4e-95;						
Matches	248;	Conservative	2;	Mismatches	0;	Indels	4;	Gaps
QY	125	LKLPDNTQTKKPEYLMKDGKSTADLTGPIILQTYRAIANYEKTSGSEMALSTGDDVVE	184					
Db	1	LKLPDNTQTKKPEYLMKDGKSTADLTGPIILQTYRAIADYEKTSGSEMALSTGDDVVE	60					
QY	185	VVEKSESQWFWCQMAKRGWIPASFLPLELSDPEDTEDPNPVAGEPYVAIKAYTAVEGDE	244					
Db	61	VVEKSESQWFWCQMAKRGWIPASFLPLELSDPEDTEDPNPVAGEPYVAIKAYTAVEGDE	120					
QY	245	VSLLLEGEAEVEIHKLLDGW-----KDDVTGYFSPMYLQKSGQDVVSQAQROIKRGAPPRSS	300					
Db	121	VSLLLEGEAEVEIHKLLDGMVIRKDDVTGYFSPMYLQKSGQDVVSQAQROIKRGAPPRSS	180					
QY	301	IRNAHSIHQSRKRLSQDAYRRNSVRFLOQRRQARPGQSPGSPGLEEEERQOTORSKPQA	360					
Db	181	IRNAHSIHQSRKRLSQDAYRRNSVRFLOQRRQARPGQSPGSPGLEEEERQOTORSKPQA	240					

RC TISSUE-BRAIN;
RA Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007878; BAA24848.1; -
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 5.
DR SMART; SM00376; SH3; 5.
SQ SEQUENCE 940 AA; 102617 MW; 1CA24D7BC03B8AC3 CRC64;

Query Match 13.5%; Score 275; DB 4; Length 940;
Best Local Similarity 26.9%; Pred. No. 2e-13;
Matches 70; Conservative 46; Mismatches 86; Indels 58; Gaps 7;

Qy 156 ILQTYRIANVETSGEMALSTGDDVVVVEKSESGWFWFQMKAKRGWIPASELEPLDS 215
Db 1 MLEQYVVVSNKKQENSELSQAGEVVDVIEKNSGWWFVSTSEEGWVPATYLEAQNG 60
Qy 216 PDETEPEPNYAG--EPVAITAKATVAGSEVSLLEGAEEVHKLKDGW-----KDD 266
Db 61 TRDSDINTSGTGEKVTVPQYTSQSKDEIGFEKGVTVVEIRKNLEGWYIRYLKRE- 119
Qy 267 VTGYFPSSYLQSGDVSSQAQRI-----KRGAPPR----- 297
Db 120 --GNAPASYLAKAKDDLPTRKKNLGAHPVEIIGNIMEISNLLNKRASGDKETPPAEGEGE 177
Qy 298 -----RSSIRNAHSIHQRKRLSQDAYRNSVRFLQRRRQARPGQSPG 343
Db 178 APIAKKEISLPILNCASNGSAGVDPDRTVSLAQS--PAVARIAPOAQISSPNLRT 235
Qy 344 SPLEERQTORSKPOPAPVP 363
Db 236 PP-RRESSLGFLPKPPPP 254

RESULT 14
Q9GMZ5
ID Q9GMZ5 PRELIMINARY; PRT; 339 AA.
AC Q9GMZ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P40-PROX.
OS Tursiops truncatus (Atlantic bottle-nosed dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Tursiops.
OX NCBI_TaxID=9739;
RN [1]
RP SEQUENCE FROM N.A.
RA Inoue Y., Itou T., Sakai T.;
RT "Molecular Cloning and Identification of Bottle-Nosed Dolphin p40-phox.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038267; BAB11805.1; -
DR InterPro; IPR000919; NCF_P40.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR000270; OPR.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00564; OPR; 1.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00497; P40PHOX.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.

SQ SEQUENCE 339 AA; 39363 MW; B6306ADD3D44E686 CRC64;

Query Match 9.7%; Score 197; DB 6; Length 339;
Best Local Similarity 25.7%; Pred. No. 8.4e-08;
Matches 83; Conservative 48; Mismatches 110; Indels 82; Gaps 13;

Qy 8 HIALLGFEKRFVPSQHYVYMFVLRWQDLSEKVVYRRFTEIYEFHKTLEKMFPIEAGAINP 67
Db 25 NIADIEKKGF--TSHFVFEVTKGSKYLIYRRYRQFVALQSKLEERF-----GP 75
Qy 68 ENR-----IIPHPAPKWDG--QRAAENROGTUTEYVCSITLSLPTKISRCPHLLD--- 116
Db 76 ENKASPYICILPTLPA-KVYVGVRQEIAEMRIPALNAYMKHLLSLPIWV-----LMDEDV 129
Qy 117 ---FEKVRPDLKLPDNDQTKKPEYLMKDGKSTADITPIILQTYRATANYEKTSGS 173
Db 130 RIFFYQSSYDAEQVPQALRRRPRT----RKVKSTSPQGFDRMAAPRAEALFDFTGNS 185
Qy 174 EMALS--TGDVVEVVEKSESGWFWFQMKAKRGWIPASELEPLDSPDETEP----- 222
Db 186 KLELNFKYGDVIFULLSRINKDWLEGTVRGTTGIFPVSEVKILKDFPEEDPTNWLRCYY 245
Qy 223 -----EPNAGEPY-----VAIRKATVAGEDEVSLLEGEAVE 254
Db 246 EDTISTIKDIAVEEDLKSTPLFKDLLELMRRREFQEDIALN-YQDAEGDLVRLLSDEVD 304
Qy 255 VIHKLKLDG-----WKDDVT 268
Db 305 LMVWQTGLPSOKHLFPWKLHIT 327

RESULT 15
Q9NZM2
ID Q9NZM2 PRELIMINARY; PRT; 1248 AA.
AC Q9NZM2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE INTERSECTIN 2 SHORT ISOFORM.
DE ITS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pucharcos C., Estivill E., de la Luna S.;
RT "Cloning and characterization of human intersectin 2: a protein involved in endocytosis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL; AF182199; AAF59904.1; -
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15_repeat.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehand; 2.
DR Pfam; PF00018; SH3; 5.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE; PS00002; SH3; 5.
KW Calcium-binding.
SQ SEQUENCE 1248 AA; 141718 MW; A325BC186C76A97D CRC64;

Query Match 9.2%; Score 188; DB 4; Length 1248;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:07:58 ; Search time 13.31 Seconds
(without alignments)
1063.308 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHIALGFEKRVFP.....ADLILNRCSESTKRLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2018	99.2	390	1 NCF1_HUMAN	P14598 homo sapien
2	1754	86.2	392	1 NCF1_BOVIN	O77774 bos taurus
3	1689	83.0	390	1 NCF1_MOUSE	Q09014 mus musculus
4	188.5	9.3	339	1 NCF4_MOUSE	P97369 mus musculus
5	172.5	8.5	339	1 NCF4_HUMAN	Q15080 homo sapien
6	157	7.7	520	1 ITSN_HUMAN	Q15811 homo sapien
7	136.5	6.7	671	1 VINE_HUMAN	O60504 homo sapien
8	135.5	6.7	733	1 VINE_MOUSE	Q07128 mus musculus
9	133.5	6.6	1634	1 PK3B_HUMAN	Q00750 homo sapien
10	133.5	6.6	2472	1 SPKN_HUMAN	Q13813 homo sapien
11	131.5	6.5	303	1 CRKL_HUMAN	P46109 homo sapien
12	131.5	6.5	2415	1 SPCA_DROME	P13395 drosophila
13	131.5	6.5	2477	1 SPCN_CHICK	P07751 gallus gall
14	129.5	6.4	303	1 CRKL_MOUSE	P47941 mus musculus
15	129.5	6.4	517	1 FGR_MOUSE	P14234 mus musculus
16	127.5	6.3	505	1 SRK1_SPOLA	P42686 spongilla l
17	127.5	6.3	1149	1 DBS_MOUSE	Q64096 mus musculus
18	127	6.2	1244	1 SLAL_YEAST	P22790 saccharomyc
19	124	6.1	305	1 CRK_CHICK	Q04929 gallus gall
20	122	6.0	1147	1 MYSB_ACACA	P19706 acanthamoeb
21	121.5	6.0	501	1 UVS2_NEUCR	P33288 neurospora
22	121	5.9	536	1 SCD2_SCHPO	P40996 schizosacch
23	121	5.9	642	1 YB65_SCHPO	O09746 schizosacch
24	119.5	5.9	359	1 YKA7_CAEEL	P34258 caenorhabdi
25	119.5	5.9	403	1 STAC_MOUSE	P97306 mus musculus
26	119	5.8	646	1 Y142_HUMAN	Q14155 homo sapien
27	116	5.7	1113	1 MYSD_DICDI	P34109 dictyosteli
28	115	5.7	381	1 NCK2_HUMAN	O43639 homo sapien
29	114.5	5.6	506	1 SRK4_SPOLA	P42690 spongilla l
30	113.5	5.6	695	1 AMPH_HUMAN	P49418 homo sapien
31	113.5	5.6	1333	1 CC25_CANAL	P43069 candida alb
32	113	5.6	343	1 SNXG_HUMAN	P57768 homo sapien
33	113	5.6	377	1 NCK1_HUMAN	P16333 homo sapien

34	113	5.6	526	1 SRC_AVISR	P00525 avian sarco
35	113	5.6	532	1 SRC_CHICK	P00523 gallus gall
36	113	5.6	557	1 SRC_AVIS1	P14085 avian sarco
37	113	5.6	568	1 SRC_AVIS2	P14084 avian sarco
38	113	5.6	587	1 SRC_AVIS2	P15054 avian sarco
39	112	5.5	304	1 CRK_MOUSE	Q64010 mus musculu
40	111.5	5.5	304	1 CRK_RAT	Q83768 rattus norv
41	111.5	5.5	402	1 STAC_HUMAN	Q99469 homo sapien
42	110	5.4	633	1 YHR4_YEAST	P38822 saccharomyc
43	109.5	5.4	296	1 CRK_XENLA	P87378 xenopus lae
44	108	5.3	304	1 CRK_HUMAN	P46108 homo sapien
45	107.5	5.3	861	1 SMAD_MOUSE	O09126 mus musculu

ALIGNMENTS

RESULT 1	
NCF1_HUMAN	
ID NCF1_HUMAN STANDARD; PRT; 390 AA.	
AC P14598;	
DT 01-APR-1990 (Rel. 14, Created)	
DT 01-MAY-1992 (Rel. 22, Last sequence update)	
DT 20-AUG-2001 (Rel. 40, Last annotation update)	
DE NEUTROPHIL CYTOSOL FACTOR 1 (NCF-1) (NEUTROPHIL NADPH OXIDASE FACTOR	
DE 1) (47 KDA NEUTROPHIL OXIDASE FACTOR) (P47-PHOX) (NCF-47K) (47 KDA	
DE AUTOSOMAL CHRONIC GRANULOMATOUS DISEASE PROTEIN).	
GN NCF1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=89332501; PubMed=2547247;	
RA Lomax K.J., Leto T.B., Nunoi H., Gallin J.I., Malech H.L.;	
RT "Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in	
RT chronic granulomatous disease.";	
RL Science 245:409-412(1989).	
RN [2]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=89386707; PubMed=2550933;	
RA Volpp B.D., Nauseef W.M., Clark R.A.;	
RT "Cloning of the cDNA and functional expression of the 47-kilodalton	
RT cytosolic component of human neutrophil respiratory burst oxidase.";	
RL Proc. Natl. Acad. Sci. U.S.A. 86:7195-7199(1989).	
RN [3]	
RP REVISIONS.	
RA Volpp B.D., Nauseef W.M., Clark R.A.;	
RL Proc. Natl. Acad. Sci. U.S.A. 86:9563-9563(1989).	
RN [4]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=90377229; PubMed=2398896;	
RA Rodaway A.R.F., Teahan C.G., Casimir C.M., Segal A.W., Bentley D.L.;	
RT "Characterization of the 47-kilodalton autosomal chronic	
RT granulomatous disease protein: tissue-specific expression and	
RT transcriptional control by retinoic acid.";	
RL Mol. Cell. Biol. 10:5388-5396(1990).	
CC -1- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE	
CC REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY	
CC FOR SUPEROXIDE PRODUCTION).	
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.	
CC -1- DISEASE: NCF1 IS ABSENT IN MOST PATIENTS WITH AUTOSOMAL RECESSIVE	
CC CHRONIC GRANULOMATOUS DISEASE (AR-CGD).	
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.	
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.	
CC -1- DATABASE: NAME=NCF1base; NOTE=NCF1 deficiency database;	
CC WWW="http://www.uta.fi/inf/bioinfo/NCF1base/"	

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way


```
Db 360 RPKQPAVPPRPSADLILHRCSESTKRKLASAV 392
RESULT 3
NCF1_MOUSE
ID NCF1_MOUSE STANDARD; PRT; 390 AA.
AC Q09014; Q07014; Q09J134;
DT 01-NOV-1995 (Rel. 32, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 1 (NCF-1) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 1) (47 KDA NEUTROPHIL OXIDASE FACTOR) (P47-PHOX) (NCF-47K).
GN NCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Macrophage;
RX MEDLINE=94164697; PubMed=8119734;
RA Jackson S.H., Malech H.L., Kozak C.A., Lomax K.J., Gallin J.I.,
RA Holland S.M.;
RT "Cloning and functional expression of the mouse homologue of
RT p47phox.";
RL Immunogenetics 39:272-275(1994);
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Green E.D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NCF2, NCF1, AND A MEMBRANE BOUND CYTOCHROME B558 ARE
CC REQUIRED FOR ACTIVATION OF THE LATENT NADPH OXIDASE (NECESSARY
CC FOR SUPEROXIDE PRODUCTION).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11455; AAA50469.1; -
DR EMBL; AB002663; BAA25649.1; -
DR EMBL; AF267747; AAF90134.1; -
DR MGD; MGI:97283; Ncf1.
DR InterPro; IPR001655; P47PHOX.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 2.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRINTS; PR00498; P47PHOX.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS50002; SH3; 2.
KW SH3 domain.
FT DOMAIN 9 128 PX.
FT DOMAIN 211 254 ASP/GLU-RICH (HIGHLY ACTIDIC).
FT DOMAIN 292 390 ARG/LYS-RICH (HIGHLY BASIC).
```

```
FT DOMAIN 162 215 SH3 1.
FT DOMAIN 226 285 SH3 2.
FT CONFLICT 161 161 Y -> H (IN REF. 1).
FT CONFLICT 343 345 GOL -> RAA (IN REF. 1).
FT CONFLICT 344 344 Q -> P (IN REF. 3).
SQ SEQUENCE 390 AA; 44698 MW; A383DB953839CFGB CRC64;

Query Match 83.0%; Score 1689; DB 1; Length 390;
Best Local Similarity 80.8%; Pred. No. 1.7e-116;
Matches 316; Conservative 34; Mismatches 35; Indels 6; Gaps 3;

Qy 1 MGDFTIRHALLGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKLKEMPEPI 60
Db 1 MGDFTIRHALLGFEKRVPSQHYVYMFVWKQDLSEKVVYRRFTEIYEFHKLKEMPEPI 60
Qy 61 EAGAINPENRIIPHPAPKWFEDGQRAENRQGLTLEYCSTLMSLPTKISRCPHLLDFFKV 120
Db 61 EAGEIHTENRVIPHPAPRWFEDGQRAESRQGLTLEYFNGLMGLPVKISRCPHLLDFFKV 120
Qy 121 RPDDLKLPDNTQTKKPEYLMKPKGCKSTATDITGPIILQTYRAIANYEKTSSENALSTG 180
Db 121 RPDDLKLPDNTQTKKPEYLMKPKGCKNNVADITGPIILQTYRAIADYEKSSGTETVATG 180
Qy 181 DVVEVVEKSESGWFCOMKAKRGWIPASFLPLEDSDPDETPENYAGEPYVAIKAYTAV 240
Db 181 DVVDVVEKSESGWFCOMKTRGWVPASVLEPLDSPDEADDPNYPAGEPYVTIKAYAAV 240
Qy 241 EGDEVSLLEGEAVEVIHKLDDGW---KDDVTGYFPSPHYLQKSGQDVSOAQRIK-RGAP 295
Db 241 EEDMSLSEGEAIEVIHKLDDGWVYVRKGDITGYSPHYLQKAGEITQAQRIKRGAP 300
Qy 296 PRSRIRNAHSTHORSKRLSDAYRRNSVRFLOQRROARPGSPGSPGLEEERQOTORS 355
Db 301 PRSTIRNAQSTHORSKRLSDQTYRRNSVRFLOQRROARPGSPGSPGLEEERQOTORS 359
Qy 356 KPQPAVPPRPSADLILHRCSESTKRKLASAV 386
Db 360 KPQPAVPPRPSADLILHRCSESTKRKLTSAV 390
```

```
RESULT 4
NCF4_MOUSE
ID NCF4_MOUSE STANDARD; PRT; 339 AA.
AC P97369;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE NEUTROPHIL CYTOSOL FACTOR 4 (NCF-4) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 4) (P40-PHOX) (P40PHOX).
GN NCF4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97152411; PubMed=8995189;
RA Zhan S., Kozak C.A., Zhan S., Chanock S.J.;
RT "Cloning and chromosomal localization of ncf4, the mouse homologue of
RT p40-phox.";
RL Immunogenetics 45:217-219(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RX MEDLINE=98149672; PubMed=9490028;
RA Mizuki K., Kadomatsu K., Hata K., Ito T., Fan Q.-W., Kage Y.,
RA Fukumaki Y., Sakaki Y., Takeshige K., Sumimoto H.;
RT "Functional modules and expression of mouse p40(phox) and p67(phox),
RT SH3-domain-containing proteins involved in the phagocyte NADPH oxidase
RT complex.";
RL Eur. J. Biochem. 251:573-582(1998).
CC -!- FUNCTION: COMPONENT OF THE NADPH-OXIDASE, A MULTICOMPONENT ENZYME
```

```
CC SYSTEM RESPONSIBLE FOR THE OXYDATIVE BURST IN WHICH ELECTRONS ARE
CC TRANSPORTED FROM NADPH TO MOLECULAR OXYGEN, GENERATING REACTIVE
CC OXIDANT INTERMEDIATES. IT MAY BE IMPORTANT FOR THE ASSEMBLY AND/OR
CC ACTIVATION OF THE NADPH-OXYDASE COMPLEX.
CC -1- SUBUNIT: P40-PHOX ASSOCIATES PRIMARILY WITH P67-PHOX TO FORM A
CC COMPLEX WITH P47-PHOX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U59488; AAC53122.1; -.
CC EMBL; AB002665; BAA25651.1; -.
CC HSSP; P29355; 2SEM.
CC MGD; MGI:109186; NCF4.
CC InterPro; IPR000919; NCF_P40.
CC InterPro; IPR000270; OPR.
CC InterPro; IPR001683; PX.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00564; OPR; 1.
CC Pfam; PF00787; PX; 1.
CC Pfam; PF00016; SH3; 1.
CC PRINTS; PR00452; SH3DOMAIN.
CC PRINTS; PR00497; P40PHOX.
CC SMART; SM00016; OPR; 1.
CC SMART; SM00312; PX; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS00002; SH3; 1.
KW SH3 domain.
FT DOMAIN 24 143 PX.
FT DOMAIN 170 229 SH3.
SQ SEQUENCE 339 AA; 38734 MW; 62C4B3252DC4B98 CRC64;

Query Match 9.3%; Score 188.5; DB 1; Length 339;
Best Local Similarity 26.2%; Pred. No. 8.2e-07;
Matches 80; Conservative 40; Mismatches 126; Indels 59; Gaps 10;

QY 15 EKRFPVSHYVMFLVKQDISEKVVYRFTYEYFHTLKMFPFIEAGINPENRIIHP 74
DB 31 EKRGFTS-HEVFIEVTKGSKYLIYRQFYALQSKLEERFGPES-KNSPFTCNLPT 88
QY 75 LPAPKWFDS-QRAAENRQGTLTGYCSTLMSLPTKISRCPHLLDFFKVRPDDLKLPDNDQT 133
DB 89 LPAPKYMGAQOEIATRIPALNAYMKNLLSLPVCVLMDDPVRIFYQSAIDA-----QV 143
QY 134 KKPETLYMPKDGKSTATDITGPIL--LQTYRAIANYEKTSGSEMAIS--TGDWVEVVEKS 189
DB 144 PQALRLRPRTKIKGVSPQGAIMDRMEAPRAEALFDFTGNSKLESLFKAGDVIFLLSKI 203
QY 190 ESGWTFQCMKAKRGWIPASFLPLDSPDETP-----EPNTAG 228
DB 204 NKDWLEGTSGQATGIFPGSFVKILKDFDEDDTNWLRICYFYEDTGKTIKIDIAVEEDLSL 263
QY 229 EPY-----VAIKATVAVGEVDSLEGEAEVVIHKLDDG-----W 263
DB 264 TPLFKDLALMRREFQREDIAL-SIQDAGDLVRLSLSDVGLVWKQARGLPQSKRLFPW 322
QY 264 KDDVT 268
DB 323 KLHVT 327

RESULT 5
NCF4_HUMAN
>ID NCF4_HUMAN STANDARD; PRT; 339 AA.
```

```
AC Q15080; O60808; Q9NP45;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUTROPHIL CYTOSOL FACTOR 4 (NCF-4) (NEUTROPHIL NADPH OXIDASE FACTOR
DE 4) (P40-PHOX) (P40PHOX).
GN NCF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94107216; PubMed=8280052;
RA Wientjes F.B., Hsuan J.J., Totty N.F., Segal A.W.;
RT "p40phox, a third cytosolic component of the activation complex of the
RT NADPH oxidase to contain src homology 3 domains.";
RL Biochem. J. 296:557-561(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96437051; PubMed=8839867;
RA Zhan S., Vazquez N., Zhan S., Wientjes F.B., Budarf M.L., Schrock E.,
RA Ried T., Green E.D., Chanock S.J.;
RT "Genomic structure, chromosomal localization, start of transcription,
RT and tissue expression of the human p40-phox, a new component of the
RT nicotinamide adenine dinucleotide phosphate-oxidase complex.";
RL Blood 88:2714-2721(1996).
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99364495; PubMed=10437784;
RA Hasebe T., Someya A., Nagaoka I.;
RT "Identification of a splice variant mRNA of p40phox, an NADPH oxidase
RT component of phagocytes.";
RL FEBS Lett. 455:257-261(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beate D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carden C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.B.,
RA Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Mathews L., Mccann O.T., Mcclay J., McLaren S., Mcmurray A.A.,
RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Willingham D.,
RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
```

RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyraud M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22";
RL Nature 402:489-495(1999).
CC -!- FUNCTION: COMPONENT OF THE NADPH-OXIDASE, A MULTICOMPONENT ENZYME
CC SYSTEM RESPONSIBLE FOR THE OXYDATIVE BURST IN WHICH ELECTRONS ARE
CC TRANSPORTED FROM NADPH TO MOLECULAR OXYGEN, GENERATING REACTIVE
CC OXIDANT INTERMEDIATES. IT MAY BE IMPORTANT FOR THE ASSEMBLY AND/OR
CC ACTIVATION OF THE NADPH-OXIDASE COMPLEX.
CC -!- SUBUNIT: P40-PHOX ASSOCIATES PRIMARILY WITH P67-PHOX TO FORM A
CC COMPLEX WITH P47-PHOX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO HEMATOPOIETIC
CC CELLS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50729; AAB39970.1; -
DR EMBL; U50720; AAB39970.1; JOINED.
DR EMBL; U50721; AAB39970.1; JOINED.
DR EMBL; U50722; AAB39970.1; JOINED.
DR EMBL; U50723; AAB39970.1; JOINED.
DR EMBL; U50724; AAB39970.1; JOINED.
DR EMBL; U50725; AAB39970.1; JOINED.
DR EMBL; U50726; AAB39970.1; JOINED.
DR EMBL; U50727; AAB39970.1; JOINED.
DR EMBL; U50728; AAB39970.1; JOINED.
DR EMBL; X77094; CAA5372.1; -
DR EMBL; AB025220; BAA89792.1; -
DR EMBL; AB025219; BAA89791.1; -
DR EMBL; AL008637; CAA15486.1; -
DR HSSP; P29355; 3SEM.
DR MIM; 601488; -
DR InterPro; IPR000919; NCF_P40.
DR InterPro; IPR000270; OPR.
DR InterPro; IPR001683; PX.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00564; OPR; 1.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00497; P40PHOX.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain; Alternative splicing.
FT DOMAIN 24 143 PX.
FT DOMAIN 170 229 SH3.
FT VARIANT 147 147 I -> L.
FT /FTID=VAR_009314.
FT DIAVEDLSPTLLKDLLELTRRREFQEDIALNVRDAGDL
FT VRLLSDEVALMVRQARGLPQSKRLFFWKLHITOKDNRYV
FT NMP -> SVAWEGACPAFLPSLRPPPLTSPSHGSLHSHK
FT APSGQSMHNAVTSQRPGWQPSPFPPTPHFPQDASL
FT LQVPTPLGTSRWKRSALPY (IN ISOFORM 2).
SQ SEQUENCE 339 AA; 39031 MW; 9A099AC0E834F90C CRC64;

Query Match 8.5%; Score 172.5; DB 1; Length 339;
Best Local Similarity 24.4%; Pred. No. 1.2e-05;
Matches 76; Conservative 48; Mismatches 116; Indels 71; Gaps 12;

QY 15 EKRFVPSOHVYVYMLVKWDLSEKVVYRRETEIYEFHKTLEKMFPIEAGAINPENR---- 70
DB 31 EKRGFTS-HFVIEVTKGSKYLIYRYRQFHALQSKLEERF-----GPDSSKSNL 82
QY 71 --IIPHLPAKWFDC--QRAAENRQGTLEYECSTLMSLPTKISRCPLHLDFEKKVRP-DDL 125
DB 83 ACTLPTLPA-KVVGKVEIAEMRIPALNAYMKSLSLSPVWVLMDEDDVRIFFYQSPYDSE 141
QY 126 KLPTDNTQTKPKETYLMPKDGKSTATDTITGPILQIYRAIANYEKTSGSEMAILS--TGOVV 183
DB 142 QVPOAIRLRPRPT----RKVKSVPQGNVSVDRAAPAEALDFDTCSKLELNFAGDVI 197
QY 184 EVVEKSESGWFWCOMKAKRGWIPASFLPLDSDPDEDP----- 222
DB 198 FLLSRINKDLGTVRGATGIFPLSFVKILKDFPEDDPTNWLRCYIYEDTISTIKDIATV 257
QY 223 EPNYAGEPY-----VAIKAYTAVGEVSLLEGEAVEVI-----H 257
DB 258 EEDLSSTPLKDLLETRRREFQEDIALN-YRDAEGDLVRLLSDEEDVALMVRQARGLPQ 316
QY 258 KLLDGWKDDVT 268
DB 317 KRLFPWKLHIT 327
RESULT 6
ITSN_HUMAN
ID ITSN_HUMAN STANDARD; PRT; 520 AA.
AC Q15811;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE INTERSECTIN (SH3 DOMAIN-CONTAINING PROTEIN SH3P17).
GN ITSN OR SH3DIA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA MEDLINE=98294438; PubMed=9630982;
RX Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
RT "Cloning of ligand targets: systematic isolation of SH3 domain-
RT containing proteins";
RL Nat. Biotechnol. 14:741-744(1996).
RN [2]
RP GENE MAPPING.
RX MEDLINE=98127038; PubMed=9465890;
RA Chen H., Antonarakis S.E.;
RT "The SH3DIA gene maps to human chromosome 21q22.1-->q22.2";
RL Cytogenet. Cell Genet. 78:213-215(1997).
CC -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U61166; AAC50592.1; -
DR HSSP; P29354; IGFDP.
DR MIM; 602442; -
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 4.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00002; SH3; 4.
KW SH3 domain; Repeat.
FT DOMAIN 116 177 SH3 1.
FT

FT DOMAIN 284 342 SH3 2.
FT DOMAIN 374 438 SH3 3.
FT DOMAIN 455 514 SH3 4.
SQ SEQUENCE 520 AA; 57976 MW; 614043FIDB098C60 CRC64;

Query Match 7.7%; Score 157; DB 1; Length 520;
Best Local Similarity 28.8%; Pred. No. 0.00029;
Matches 47; Conservative 35; Mismatches 61; Indels 20; Gaps 9;

QY 142 PKDGTATDTGPIIL--QYTRATANTKSGSEMASTGDVVEVVKSGWFWCOMK 199
DB 358 PASLRVASPAKPVYSGEIAQVATATGPEQLTAPGLILIRKKNPGWMEGLQ 417
QY 200 A---KR--GWIPASFLPLDSP-----DETEDEPNYAGE--PVVAIKAYTAVEGDEVSL 247
DB 418 ARGKRQGWFPANVKKL--SPGTSKIYTPTEPKSTALAAVCQVIGMDYTAONDELAF 476
QY 248 LEGEAVEVIHKL-LDGWKDDV---TGYFSPMYLQ-KSGQDVQSQ 285
DB 477 NKGQIINVLNKEDPDWKGVEYNGVGLFPSNVYKLTITMDPFSQ 519

RESULT 7

VINE_HUMAN
ID VINE_HUMAN STANDARD; PRT; 671 AA.
AC O60504; OSUQEA;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VINEXIN (SH3-CONTAINING ADAPTOR MOLECULE-1) (SCAM-1).
GN SCAM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS.
RC Tissue=Placenta;
RX MEDLINE=99102423; PubMed=9885244;
RA Kioka N., Sakata S., Kawachi T., Amachi T., Akiyama S.K., Okazaki K.,
RA Yaen C., Yamada K.M., Aota S.-I.;
RT "Vinexin: a novel vinculin-binding protein with multiple SH3 domains
RT enhances actin cytoskeletal organization.";
RL J. Cell Biol. 144:59-69(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RA Her J.-H., Gorman D., Miyajima A., Bolen J.B.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH SOS.
RX MEDLINE=20054482; PubMed=10585480;
RA Akamatsu M., Aota S.-I., Suwa A., Ueda K., Amachi T., Yamada K.M.,
RA Akiyama S.K., Kioka N.;
RT "Vinexin forms a signaling complex with Sos and modulates epidermal
RT growth factor-induced c-Jun N-terminal kinase/stress-activated
RT protein kinase activities.";
RL J. Biol. Chem. 274:35933-35937(1999).
CC -!- FUNCTION: VINEXIN ALPHA ISOFORM PROMOTES UPREGULATION OF ACTIN
CC STRESS FIBER FORMATION. VINEXIN BETA ISOFORM PLAYS A ROLE IN CELL
CC SPREADING AND ENHANCES THE ACTIVATION OF JNK/SAPK IN RESPONSE TO
CC EGF STIMULATION BY USING ITS THIRD SH3 DOMAIN.
CC -!- SUBUNIT: INTERACTS WITH VINCULIN BY THE FIRST TWO SH3 DOMAINS AND
CC THE PROLINE RICH REGION OF VINCULIN. BINDS TO SOS (GUANINE
CC NUCLEOTIDE EXCHANGE FACTOR OF RAS AND RAC), THROUGH ITS THIRD SH3
CC DOMAIN. THE FORMATION OF THIS COMPLEX IS DOWN-REGULATED BY
CC PHOSPHORYLATION OF SOS.
CC -!- SUBCELLULAR LOCATION: BOTH ISOFORMS WERE LOCALIZED AT FOCAL
CC ADHESION AND CELL-CELL ADHESIONS SITES, VINEXIN BETA WAS ALSO
CC FOUND IN THE NUCLEUS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: BOTH ISOFORMS ARE EXPRESSED IN DIFFERENT

CC TISSUES LIKE HEART, PLACENTA, BRAIN, SKELETAL MUSCLE AND
CC PRANCREAS. VINEXIN BETA ISOFORM IS ESPECIALLY FOUND IN LIVER.
CC -!- SIMILARITY: CONTAINS 3 SH3 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF064807; AAD32304.1; -;
DR EMBL; AF037261; AAC09244.1; -;
DR HSSP; P07751; 1BK2.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR003127; Sorb.
DR Pfam; PF00018; SH3; 3.
DR Pfam; PF02208; Sorb; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00326; SH3; 3.
DR SMART; SM00459; Sorb; 1.
DR PROSITE; PS50002; SH3; 3.
KW Cytoskeleton; SH3 domain; Repeat; Cell adhesion; Alternative splicing.
FT DOMAIN 117 164 SORB.
FT DOMAIN 380 439 SH3 1.
FT DOMAIN 454 515 SH3 2.
FT DOMAIN 612 671 SH3 3.
FT DOMAIN 380 515 BINDS TO VINCULIN.
FT DOMAIN 612 671 BINDS TO SOS.
FT VARSPLIC 1 342 MISSING (IN ISOFORM BETA).
FT MUTAGEN 649 649 W->E: LOSS OF SOS-BINDING ABILITY.
FT MUTAGEN 667 667 Y->V: LOSS OF SOS-BINDING ABILITY.
FT CONFLICT 583 583 L -> F (IN REF. 1).
SQ SEQUENCE 671 AA; 75328 MW; 7D978711726BCD32 CRC64;

Query Match 6.7%; Score 136.5; DB 1; Length 671;
Best Local Similarity 22.6%; Pred. No. 0.013;
Matches 88; Conservative 51; Mismatches 140; Indels 11; Gaps 19;

QY 46 EYEPFKTL-REMFPTEAGAINPENRIIPAPKWFQDQRAAENR--QGTLTEYCSILM 102
DB 278 ELAELSAELDKDLRAITRLPSKSS-----PAPR-----RAPEORPPAGPASWSSVP 327
QY 103 SLP---TKISCPHLL-----DFFKVRPDDLKLPDNTQTKKETIYMPKDGKST 148
DB 328 HAPYLGARSLSPHKMDGGSPFGLGRDF--VYPSSTRDPSASNGGSPARREKKRKA 385
QY 149 ATDITGPILOTYRAIANVKTSGSEMASTGDDVVVEVVKSGWFWCOMKAKRWIPAS 208
DB 386 -----LUKDFDQAQSPKELTQKGDIVYIHKVDKKNLGEHIGRLGIFPAN 432
QY 209 FLEPLDSPDETDP--EPNYAGEPVVAIKAYTAVBGD---EVSLEGEAVEVIHKLDDGW 263
DB 433 YVEVLPA-DEIPKPIKPTTYQVLEYGEAVQYTFKGDLEVELSPFKGHEICLRKVNENW 491
QY 264 KDD-VT-----GYFSPMYLQKSGQDVSOAQKIKGAPRRSSIRNAHSIHQSRKRLSQ 317
DB 492 YEGRITGTGROGIFPASYVQ-----VSREPRLRCD 522
QY 318 DAYRRNSVRELQORROARPGQSPGS-----PLEEER-OTQ 353
DB 523 DGQPLTSPRLTAARSAR--HPSSPSALRSADPTDLGGQTSPRKTGTFPTQEPPTQ 581
QY 354 R-SKQPAPVPPR--PSADLIILNRCSESTKR 380
DB 582 NLGTGPPALSHRGSFHPDLGLTSSPNTSQ 611

RESULT 8
VINE_MOUSE

DR Pfam; PF00435; spectrin; 23.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00150; SPEC; 20.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00002; SH3; 1.
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;
KW Capping protein; Calcium-binding; Duplication; Repeat; SH3 domain.
FT DOMAIN 1 14
FT REPEAT 15 119
FT REPEAT 120 225
FT REPEAT 226 331
FT REPEAT 332 437
FT REPEAT 438 543
FT REPEAT 544 648
FT REPEAT 649 754
FT REPEAT 755 860
FT REPEAT 861 966
FT DOMAIN 967 1061
FT DOMAIN 967 1026
FT DOMAIN 1062 2251
FT REPEAT 1062 1167
FT REPEAT 1204 1309
FT REPEAT 1310 1415
FT REPEAT 1416 1521
FT REPEAT 1522 1628
FT REPEAT 1629 1734
FT REPEAT 1735 1840
FT REPEAT 1841 1946
FT REPEAT 1947 2053
FT REPEAT 2054 2166
FT REPEAT 2167 2251
FT DOMAIN 2252 2472
FT CA_BIND 2336 2347
FT CA_BIND 2379 2390
FT CONFLICT 1595 1595
SQ SEQUENCE 2472 AA; 284279 MW; CDA1964462A7028D CRC64;

Query Match 6.6%; Score 133.5; DB 1; Length 2472;
Best Local Similarity 19.3%; Pred. No. 0.12;
Matches 49; Conservative 45; Mismatches 91; Indels 69; Gaps 8;

QY 31 KQDLSEKVVYR-----FTEIYEFHKTLEMPFIEAGAINPENRIIHLPA 78
DB 871 KWEALKASORQDLSLOAQYFADANAEWSWREKEPIVGSTYCKDE----- 922
QY 79 KWFQGRAA-----ENRQGTLTETCYSLMPLTKISRCPHLLDFFKVRPDDLKLPDNTQK 134
DB 923 ---DSAEALLKHEALMSDLSAYGSSIOALREQAQCRCQV-----APTDDET 968
QY 135 KPEYLMKPKGKSTYDTGPIILQTYRAIANEYKTSSEMALSTGDVVEVVEKSESGW 194
DB 969 K-----ELVLALY----DYEKSPREVTMKKGDIILTLNNTKDW 1005
QY 195 FCOMKAKRGWIPASLEPLDSDPEDEPNYAGEPYVAIK-----AYTAVEGEVSL 247
DB 1006 KVEVNDROGFVPAVVKLLD-PAQSASRENLEEGSIALROEQIDNTRITKAGSVSL 1064
QY 248 LEAGEAVEVHKLLD 261
DB 1065 RMQVVEYHLSLE 1078

RESULT 11
CRKL_HUMAN
ID CRKL_HUMAN STANDARD; PRT; 303 AA.
AC P46109;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CRK-LIKE PROTEIN.

GN CRKL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Splice;
RX MEDLINE=93368949; PubMed=8361759;
RA ten Hoeve J., Morris C., Heisterkamp N., Groffen J.;
RT "Isolation and chromosomal localization of CRKL, a human crk-like
gene.";
RL Oncogene 8:2469-2474(1993).
CC -!- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59656; CAA42199.1; -.
DR HSSP; O64010; 1CKB.
DR MIN; 602007; -.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 2.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH2 domain; SH3 domain.
FT DOMAIN 14 102
FT DOMAIN 123 183
FT DOMAIN 235 296
SQ SEQUENCE 303 AA; 33777 MW; 294CFLEE2CD44B81 CRC64;

Query Match 6.5%; Score 131.5; DB 1; Length 303;
Best Local Similarity 21.6%; Pred. No. 0.01;
Matches 50; Conservative 36; Mismatches 85; Indels 61; Gaps 9;

QY 103 SLPTK-----ISRCPHLLDFFKVRPDDLKLPDNTQKPKETVLMKPKGKSTATDI-T 153
DB 65 SLPNRRFKIGDOEDFHLFALLEFYKIH----YLDTTTLIEPAPRYPSPPMGVSVAFLPT 120
QY 154 GPIILQTYRAIANEYKTSSEMALSTGDVVEVVEKSESGWFCOMKAKR-GWIPASFLEP 212
DB 121 AEDNLEYVRTLVDYFGNDAEDLPFKKGEILVIEKPEEQWMSARNKDGVRGMPVYVEK 180
QY 213 L--DSP-----DETEDEPNYAGEPYVAIKAYTAVEGEV-- 245
DB 181 LVRRSPHGKHGNNRNSNYGIPEPAHAYAQPTTTLPAVSGSPGAAITPLPSTQNGPVFA 240
QY 246 -----SILGEAVEVVIHKLDDG-WKDDVT---GVFPSPMYLQ 277
DB 241 KAIQKRVPCAYDKTALALEVGDIVKVTNRNNGQWEGEVNGRKGFLPPFTHVK 292

RESULT 12
SPCA_DROME
ID SPCA_DROME STANDARD; PRT; 2415 AA.
AC P13395; Q9M085; Q26340;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)


```
FT REPEAT 1044 1151 SPECTRIN 11.
FT REPEAT 1152 1257 SPECTRIN 12.
FT REPEAT 1258 1363 SPECTRIN 13.
FT REPEAT 1364 1469 SPECTRIN 14.
FT REPEAT 1470 1576 SPECTRIN 15.
FT REPEAT 1577 1682 SPECTRIN 16.
FT REPEAT 1683 1788 SPECTRIN 17.
FT REPEAT 1789 1894 SPECTRIN 18.
FT REPEAT 1895 2001 SPECTRIN 19.
FT REPEAT 2002 2115 SPECTRIN 20.
FT REPEAT 2116 2229 SPECTRIN 21.
FT REPEAT 2230 2335 SPECTRIN 22.
FT REPEAT 2336 2441 SPECTRIN 23.
FT CA_BIND 2278 2389 EF-HAND 1 (POTENTIAL).
FT CA_BIND 2321 2332 EF-HAND 2 (POTENTIAL).
FT DOMAIN 970 1029 SH3.
FT CONFLICT 110 110 Q -> D (IN REF. 3).
FT CONFLICT 1668 1668 Q -> R (IN REF. 1).
SQ SEQUENCE 2415 AA; 278301 MW; F1F72FB990EB0A37 CRC64;

Query Match 6.5%; Score 131.5; DB 1; Length 2415;
Best Local Similarity 15.3%; Pred. No. 0.16; Indels 133; Gaps 10;
Matches 56; Conservative 62; Mismatches 114;

Qy 31 KWQDLSEKVVYRR-----FTIEYEFHKLKEMFPIEAGAINPENRIIPLHPAP 78
Db 874 QWNTLKEKSSQKDLDDSLQAHQYFADANEAEWMEKEPIATSGDYKDE----- 925
Qy 79 KWFDCQRAENRQGLTE-----YCTLSMLPTKISRCPHLLDFFKVRPDDULKLPDNTQK 134
Db 926 ---DSSEALLKHEALVSDLEAFNGTIALQALQEQKNC----- 959
Qy 135 KPETYLMPKDGKSTADITGPILITLYRAIANYEKTSGEWALSTGDVVEVVEKSESGW 194
Db 960 -----RQQTTPVVDITGK---ECVVALDYTKSPREVSMMKKGDVITLNSNKNQW 1008
Qy 195 FCOMKAKRGWIPASLEPLDS-----PDTEDEPEPNYAGPEY----- 231
Db 1009 KVEVNDROGFVPAAYIKKIDAGLSAQSNLVNDSIAKRNQINSQYDNLALARERQNK 1068
Qy 232 --VAIKATAVE-----GDEVSLLGEAVEVHKLDDGKDDV 267
Db 1069 LNETVKAYVLVREAADLAQWIRDKENHQAQIDVVGEDL-----EEVEVLQKFFDNDL 1123
Qy 268 TG-----YPSMWLYKSGQDVSOAQROIKRGAPRRSSIRNAHSI 307
Db 1124 KANEVRLANMNEIAVOLTSLGQTEALKIQTOMQDLNEKWNINLQTLTAEKASQLGSAHEV 1183
Qy 308 HORSR 312
Db 1184 ORFHR 1188

RESULT 13
ID SPCN_CHICK STANDARD; PRT; 2477 AA.
AC P07751;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPECTRIN ALPHA CHAIN, BRAIN (SPECTRIN, NON-ERYTHROID ALPHA CHAIN)
DE (FODRIN ALPHA CHAIN).
GN SPTAN1 OR SPTA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89093238; PubMed=2910879;
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm M.,
RA Lehto V.-P.;
```

```
RT "Primary structure of the brain alpha-spectrin.";
RL J. Cell Biol. 108:79-93(1989).
RN [2]
RP REVISIONS.
RA Wasenius V.-M., Saraste M., Salven P., Eraama M., Holm L.,
RA Lehto V.-P.;
RL J. Cell Biol. 108:1177-1178(1989).
RN [3]
RP SEQUENCE OF 1695-2153 FROM N.A.
MEDLINE=85284928; PubMed=4029118;
RA Wasenius V.-M., Saraste M., Knowles J., Virtanen I., Lehto V.-P.;
RT "Sequencing of the chicken non-erythroid spectrin cDNA reveals an
internal repetitive structure homologous to the human erythrocyte
spectrin.";
RL EMBO J. 4:1425-1430(1985).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 965-1025.
RX MEDLINE=93063299; PubMed=1279434;
RA Musacchio A., Noble M., Paupit R., Wierenga R., Saraste M.;
RT "Crystal structure of a Src-homology 3 (SH3) domain.";
RL Nature 359:851-855(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 969-1025.
RX MEDLINE=98363217; PubMed=9699637;
RA Martinez J.C., Pisabarro M.T., Serrano L.;
RT "Obligatory steps in protein folding and the conformational diversity
of the transition state.";
RL Nat. Struct. Biol. 5:721-729(1998).
RN [6]
RP STRUCTURE BY NMR OF 1763-1872.
RX MEDLINE=98022917; PubMed=9356261;
RA Pascual J., Pfuhl M., Walther D., Saraste M., Nilges M.;
RT "Solution structure of the spectrin repeat: a left-handed
antiparallel triple-helical coiled-coil.";
RL J. Mol. Biol. 273:740-751(1997).
RN [7]
RP STRUCTURE BY NMR OF 2320-2403.
RX MEDLINE=96067121; PubMed=7588621;
RA Trave G., Lacombe J.-P., Pfuhl M., Saraste M., Pastore A.;
RT "Molecular mechanism of the calcium-induced conformational change in
the spectrin EF-hands.";
RL EMBO J. 14:4922-4931(1995).
CC -!- FUNCTION: MORPHOLOGICALLY, SPECTRIN-LIKE PROTEINS APPEAR TO BE
RELATED TO SPECTRIN, SHOWING A FLEXIBLE ROD-LIKE STRUCTURE.
THEY CAN BIND ACTIN BUT SEEM TO DIFFER IN THEIR CALMODULIN-
BINDING ACTIVITY. IN NONERYTHROID TISSUES, SPECTRINS. IN
ASSOCIATION WITH SOME OTHER PROTEINS, MAY PLAY AN IMPORTANT
ROLE IN MEMBRANE ORGANIZATION.
CC -!- SUBUNIT: LIKE ERYTHROCYTE SPECTRIN, THE SPECTRIN-LIKE PROTEINS
ARE CAPABLE TO FORM DIMERS WHICH CAN FURTHER ASSOCIATE TO
TETRAMERS.
CC -!- DOMAIN: SPECTRIN-LIKE PROTEINS HAVE FIVE DOMAINS:
(1) N-TERMINAL DOMAIN (N),
(2) DOMAIN BETWEEN THE N-TERMINAL AND MIDDLE DOMAIN (NM),
(3) MIDDLE DOMAIN (M),
(4) DOMAIN BETWEEN THE MIDDLE AND C-TERMINAL DOMAIN (MC),
(5) C-TERMINAL DOMAIN (C).
CC NM AND MC DOMAINS ARE COMPOSED OF TYPICAL SPECTRIN 106 RESIDUES
REPEATS (1-8 FOR NM AND 12-19 FOR MC) AND ARE HOMOLOGOUS TO EACH
OTHER. N, M, AND C DOMAINS ARE COMPOSED OF SEQUENCES THAT DO NOT
FORM TYPICAL SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; X14518; CAA32662.1; -;
```

```
DR EMBL; X14519; CAA32663.1; ALT_SEQ.  
DR EMBL; X02593; CAB51571.1; ALT_SEQ.  
DR PIR; A30122; SJCHA.  
DR PDB; 1SHG; 31-OCT-93.  
DR PDB; 1AEY; 15-MAY-97.  
DR PDB; 1AJ3; 07-JUL-97.  
DR PDB; 1TUC; 01-AUG-96.  
DR PDB; 1TUD; 01-AUG-96.  
DR PDB; 1BK2; 16-FEB-99.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001452; SH3.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00036; ehand; 2.  
DR Pfam; PF00018; SH3; 1.  
DR Pfam; PF00435; spectrin; 23.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR SMART; SM00054; EFH; 2.  
DR SMART; SM00326; SH3; 1.  
DR SMART; SM00150; SPEC; 20.  
DR PROSITE; PS00018; EF_HAND; 2.  
DR PROSITE; PS00002; SH3; 1.  
KW Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding;  
KW Capping protein; Calcium-binding; Duplication; Repeat; SH3 domain;  
KW 3D-structure.  
FT DOMAIN 1 14 N-TERMINAL DOMAIN.  
FT REPEAT 15 119 1.  
FT REPEAT 120 225 2.  
FT REPEAT 226 331 3.  
FT REPEAT 332 437 4.  
FT REPEAT 438 543 5.  
FT REPEAT 544 648 6.  
FT REPEAT 649 754 7.  
FT REPEAT 755 860 8.  
FT REPEAT 861 966 9.  
FT DOMAIN 967 1061 MIDDLE DOMAIN.  
FT DOMAIN 967 1026 SH3.  
FT DOMAIN 1052 2256 MC DOMAIN.  
FT REPEAT 1062 1167 11.  
FT REPEAT 1204 1309 12.  
FT REPEAT 1310 1415 13.  
FT REPEAT 1416 1521 14.  
FT REPEAT 1522 1633 15.  
FT REPEAT 1634 1739 16.  
FT REPEAT 1740 1845 17.  
FT REPEAT 1846 1951 18.  
FT REPEAT 1952 2058 19.  
FT REPEAT 2059 2171 20.  
FT REPEAT 2172 2256 21.  
FT DOMAIN 2257 2477 C-TERMINAL DOMAIN.  
FT CA_BIND 2341 2352 EF-HAND 1 (POTENTIAL).  
FT CA_BIND 2384 2395 EF-HAND 2 (POTENTIAL).  
FT STRAND 271 374  
FT STRAND 978 978  
FT TURN 983 984  
FT STRAND 985 985  
FT STRAND 988 988  
FT TURN 990 991  
FT STRAND 993 998  
FT STRAND 1004 1009  
FT TURN 1010 1011  
FT STRAND 1012 1017  
FT HELIX 1018 1020  
FT STRAND 1021 1024  
SQ SEQUENCE 2477 AA; 285361 MW; AD4C876994E6AB39 CRC64;  
  
Query Match 6.5%; Score 131.5; DB 1; Length 2477;  
Best Local Similarity 19.3%; Pred. No. 0.16;  
Matches 49; Conservative 44; Mismatches 92; Indels 69; Gaps 8;  
  
QY 31 KWQDLSEKVVYRR-----FTETPEPHKYLKEMFPTEAGAINPENRIIPLPAP 78  
QD 871 KWDSLKASQRQDLESLQAQQYFADANEQAQSWREKEPIVGSTDYGKDE----- 922
```

```
QY 79 KWFDGORA-----ENRQGTLEYVCSTLMSLPKISRCPHLLDFKVRPDDLKLPDNDQTK 134  
DB 923 ----DSAEALLKHEALMSDLISAYSGSIOALRQAQSCRQQV-----APDDETG 968  
QY 135 KPETYLMPKDGKSTATDITGPILILQTYRAIANYEKTSSEMASTGDDVVVVVEKSESGMW 194  
DB 969 K-----ELVLALY-----DYQEKSPREVTMKKGDIITLLNSINKDWW 1005  
QY 195 FCOMKAKRGWIPASFLPDLSPDETEDEPNVAGPYVAIK-----AYTAVEGDEVSL 247  
DB 1006 KVEVNDROGFVPAAYVKKID-PAQSASRENLEQGSIALRQEQIDNQTTLITKEVGSVSL 1064  
QY 248 LGEAVEVIHKLLD 261  
DB 1065 RMKQVEELYHSLLE 1078  
  
RESULT 14  
CRKL_MOUSE  
ID CRKL_MOUSE STANDARD; PRT; 303 AA.  
AC P47941;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CRK-LIKE PROTEIN.  
GN CRKL OR CRKOL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL X CHA; TISSUE=Placenta;  
RX MEDLINE=96038874; PubMed=7478571;  
RA de Jong R.L., Haataja L., Voncken J.W., Heisterkamp N., Groffen J.;  
RT "Tyrosine phosphorylation of murine Crkl.";  
RL Oncogene 11:1469-1474(1995).  
CC -!- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.  
CC -!- PTM: PHOSPHORYLATED ON TYROSINE. PHOSPHORYLATION IS PROMINENT  
CC DURING EARLY DEVELOPMENT, BUT DECREASES AT LATER EMBRYONIC STAGES  
CC AND IN NEWBORN MICE.  
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; X90648; CAA62220.1;  
DR HSSP; Q64010; 1CKB.  
DR MGD; MGI:104686; Crkol.  
DR InterPro; IPR000980; SH2.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00017; SH2; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00401; SH2DOMAIN.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR SMART; SM00252; SH2; 1.  
DR SMART; SM00326; SH3; 2.  
DR PROSITE; PS00001; SH2; 1.  
DR PROSITE; PS00002; SH3; 1.  
KW SH2 domain; SH3 domain; Phosphorylation.  
FT DOMAIN 14 102 SH2  
FT DOMAIN 123 183 SH3 1.  
FT DOMAIN 235 296 SH3 2.  
SQ SEQUENCE 303 AA; 33817 MW; B8126DDFF2053573 CRC64;
```

```
Query Match      6.4%; Score 129.5; DB 1; Length 303;
Best Local Similarity 21.6%; Pred. No. 0.015;
Matches 50; Conservative 35; Mismatches 86; Indels 61; Gaps 9;

QY 103 SLPTK-----ISRCPHLLDFKVRPDDLKLPDNTQTKKPEYLMKPKDGKSTATDI-T 153
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
Db 65 SLPNRFRKIGDQDFHLPALLEFYKIH-----YLDTTTLEPAPRYPPSPVPGSVSAPNLP 120
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
QY 154 GPIILQTVRAIANYEKTSSEENALSTGDDVVEVEKESGWFCEOMKAKR-GWIPASFLPEP 212
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
Db 121 AEENLEYVRLTYDFPGNDAEDLPFKKGELLVIEKPEQWWSARKDGRVGMIPVYVEK 180
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
QY 213 L--DSP-----DETEDEPNYAGPYPYVAIKAYTAVEGDEV-- 245
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
Db 181 LVRSSPHGKHGNNNSYGIPEAHAYAQPTTTLPTVASTPGAALNPLPTQNGPVFA 240
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
QY 246 -----SLEGEAVEVIHKLDG-WKDDVT---GYPPSWYLO 277
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |
Db 241 KAIQKRVPCAYDKTALAEVGDIVKTRMNINGOWEGVNGRKLFPFTHVK 292
    |||  | |||:|:|  | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
FGR_MOUSE
ID FGR_MOUSE STANDARD; PRT; 517 AA.
AC P14234; Q61404;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FGR (EC 2.7.1.112) (P55-FGR)
DE (C-FGR).
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2J;
RX MEDLINE=89385605; PubMed=2674853;
RA Yi T.L., Willman C.L.;
RT "Cloning of the murine c-fgr proto-oncogene cDNA and induction of
RT c-fgr expression by proliferation and activation factors in normal
RT bone marrow-derived monocytic cells.";
RL Oncogene 4:1081-1087(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BA15/C; TISSUE=Monocytic leukemia;
RX MEDLINE=90191719; PubMed=2179817;
RA King F.J., Cole M.D.;
RT "Molecular cloning and sequencing of the murine c-fgr gene.";
RL Oncogene 5:337-344(1990).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16440; CAA34463.1; -
CC EMBL; X52191; CAA36437.1; -
CC PIR; A33127; A33127.
CC PIR; S10072; S10072.
CC HSP; P00523; 2PTK.
CC MGD; MGI:95527; Fgr.
CC InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001245; Tyr_kin.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
KW Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
KW Phosphorylation; SH2 domain; SH3 domain.
FT DOMAIN 65 126 SH3.
FT DOMAIN 132 229 SH2.
FT DOMAIN 251 504 PROTEIN KINASE.
FT NP_BIND 257 265 ATP (BY SIMILARITY).
FT BINDING 279 279 ATP (BY SIMILARITY).
FT ACT_SITE 370 370 BY SIMILARITY.
FT MOD_RES 400 400 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 41 41 N -> T (IN REF. 2).
FT CONFLICT 212 212 Q -> R (IN REF. 2).
SQ SEQUENCE 517 AA; 58867 MW; F635BDB4510F3076 CRC64;

Query Match      6.4%; Score 129.5; DB 1; Length 517;
Best Local Similarity 21.0%; Pred. No. 0.029;
Matches 65; Conservative 50; Mismatches 92; Indels 103; Gaps 16;

QY 119 KVRP---DDLKLPD---NOTKKPETYLMKPKDGKSTAT--DITGPILQI----- 160
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 KLEPASKEDVGLGDFRSQTAERYYPDPTQGRNSSVFPQPTSPAFNLTNMRSISGTGV 67
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 --YRAITANYEKTSGSEMALSTGDDVVEVEKESGWFCEQ--MKAKRGWIPASFLPLDSP 216
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 TIFVALDYEARATGDDLTFTKGEKPHILNNTYDWMWEARSLSSGHRGYVPSNVVAPVDSI 127
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 217 DETE-----DPEPNY--AGEPY-----VAIKAYTAVEGDEVSLLE 249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 QAEEWYFGKISRKDAERQLLSGNFQGAFLIRESETTKGAYSLIRDWQNRGDHIKHYK 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 250 GEAVEVIHKLDGKKDDVTGYFPFSMYLQ-KSGQDVSAQQRQIKRG-----APPRSSIR 302
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 -----IRKL-----DTGGYVITTRAQFDSIQDLVQHYMEVNDGLCYLLTAPCTTTKPQ 235
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 NAHSHQSRKRRLSQDAYR--RNSVRFLOQRR-----RQARPG 338
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 TL-----GLAKDAWEIDRNSIAL--ERRLGTGCGFDGVDWLTGWCSTKVAVKTLKPG 284
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 339 POSPGSPLEE 348
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 285 TMSPKAFLEE 294
```

Search completed: January 11, 2002, 09:11:26
Job time: 208 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:05:54 ; Search time 15.91 seconds
(without alignments)
1648.107 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDFTIRHALLGFKRFPV.....ADLILNRCSESTRKRLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2018	99.2	390	1 A39249	neutrophil cytosol
2	1666	81.9	388	2 I54525	leukemia-related p
3	275	13.5	940	2 T00056	hypothetical prote
4	179.5	8.8	1270	2 T09194	adaptor protein in
5	169.5	8.3	339	2 S39768	neutrophil cytosol
6	164.5	8.1	639	2 T13151	adapter protein CM
7	164	8.1	1011	2 T13055	dynamain associated
8	161	7.9	1094	2 T13053	dynamain associated
9	150.5	7.4	665	2 JC7191	85K c-Cbl-interact
10	147.5	7.2	290	2 T42526	hypothetical prote
11	145	7.1	1119	2 T50995	related to cytoske
12	142.5	7.0	1097	2 T31504	hypothetical prote
13	139	6.8	443	2 T27877	hypothetical prote
14	136	6.7	50	4 A58224	neutrophil cytosol
15	135.5	6.7	816	2 T17257	hypothetical prote
16	133.5	6.6	1634	1 JC5500	phosphoinositide 3
17	133.5	6.6	2472	2 A35715	fodrin alpha chain
18	133	6.5	408	2 T26940	hypothetical prote
19	131.5	6.5	303	2 S41754	CRKL protein - hum
20	131.5	6.5	2415	1 A37333	spectrin alpha cha
21	131.5	6.5	2477	1 S2CHA	spectrin alpha cha
22	130.5	6.4	1196	2 T41108	SH2-containing pro
23	129.5	6.4	303	2 S58352	SH2/SH3 adaptor pr
24	127.5	6.3	505	1 S24550	protein-tyrosine k
25	127.5	6.3	1249	2 A56511	myosin I myoA - Em
26	127	6.2	1244	2 S25327	cytoskeleton assem
27	125.5	6.2	517	2 A43807	protein-tyrosine k
28	125.5	6.2	1109	2 S53601	myosin-IC [similar
29	124	6.1	305	1 A49011	c-Crk - chicken

30	124	6.1	1217	2 T39427	probable myosin I
31	123.5	6.1	648	2 T33339	hypothetical prote
32	123.5	6.0	1107	1 S52517	myosin I heavy cha
33	122	6.0	1147	1 MWX1B	myosin heavy chain
34	121.5	6.0	501	2 S34825	uvs2 protein - Neu
35	121	5.9	536	2 T38210	scd2 protein - fls
36	121	5.9	642	2 T39376	hypothetical DAG-b
37	121	5.9	728	2 T26607	hypothetical prote
38	120.5	5.9	2427	2 T16613	hypothetical prote
39	119.5	5.9	359	2 S27788	neutrophil oxidase
40	119.5	5.9	403	2 JC5269	neuron-specific si
41	118.5	5.8	408	2 T42650	hypothetical prote
42	117.5	5.8	1658	2 T42642	phosphoinositide 3
43	116.5	5.7	496	2 T46356	hypothetical prote
44	116	5.7	892	2 T09071	SH3 domains-contai
45	116	5.7	1113	1 A47106	myosin heavy chain

ALIGNMENTS

RESULT 1
A39249
neutrophil cytosol factor 1 - human
N:Alternate names: 47K autosomal chronic granulomatous disease protein; multicomponent
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence-revision 30-Jun-1993 #text-change 18-Jun-1999
C:Accession: A35926; A39249; A54067; I59190; A32762; A41385
R:Rodaway, A.R.F.; Teahan, C.G.; Casimir, C.M.; Segal, A.W.; Bentley, D.L.
Mol. Cell. Biol. 10, 5388-5396, 1990
A:Title: Characterization of the 47-kilodalton autosomal chronic granulomatous diseases
A:Reference number: A35926; MUID:90377229
A:Accession: A35926
A:Molecule type: mRNA
A:Residues: 1-390 <ROD>
A:Cross-references: GB:M55067; GB:M38755; NID:q189050; PIDN:AAA59901.1; PID:g189051
R:Volpp, B.D.; Nauseef, W.M.; Donelson, J.E.; Moser, D.R.; Clark, R.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9563, 1989
A:Reference number: A39249
A:Contents: erratum
A:Accession: A39249
A:Molecule type: mRNA
A:Residues: 1-390 <VOI>
A:Cross-references: GB:M25665; GB:M26193; NID:q189107; PIDN:AAA57209.1; PID:g189108
R:Volpp, B.D.; Nauseef, W.M.; Donelson, J.E.; Moser, D.R.; Clark, R.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 7195-7199, 1989
A:Title: Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic
A:Reference number: A41385; MUID:89386707
A:Contents: annotation
A:Note: the sequence reported has been extensively revised in reference A39249
R:Lomax, K.J.; Leto, T.L.; Nunoi, H.; Gallin, J.I.; Malech, H.L.
Science 245, 409-412, 1989
A:Title: Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in chronic g
A:Reference number: A32762; MUID:89332501
A:Contents: annotation
A:Note: the sequence reported has been extensively revised and now agrees with that s
R:Finan, P.; Shimizu, Y.; Gout, I.; Hsuan, J.; Truong, O.; Bucher, C.; Bennett, P.;
J. Biol. Chem. 269, 13752-13755, 1994
A:Title: An SH3 domain and proline-rich sequence mediate an interaction between two c
A:Reference number: A54067; MUID:94245680
A:Accession: A54067
A:Molecule type: protein
A:Residues: 8-16;44-52;71-77 <FIN>
A:Experimental source: differentiated HL-60 cells
R:Casimir, C.M.; Bu-Gharios, H.N.; Rodaway, A.R.F.; Bentley, D.L.; Rowe, P.; Segal, A.
Proc. Natl. Acad. Sci. U.S.A. 88, 2753-2757, 1991
A:Title: Autosomal recessive chronic granulomatous disease caused by deletion at a di
A:Reference number: I59190; MUID:91187670
A:Accession: I59190
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 14-24 <CAS>
A:Cross-references: GB:M60941; NID:g189948; PIDN:AAA60086.1; PID:g189949

A; Note: a GT dinucleotide deletion at a GT-GT repeat causes a frameshift after residue 222, which results in a premature stop codon and a truncated protein.
C; Comment: This protein is required for activation of the latent NADPH oxidase, which is involved in the pathogenesis of the severe combined immunodeficiency disease.
C; Genetics: Autosomal recessive.
A; Gene: GDB: NCF1
A; Cross-references: GDB: 120222; OMIM: 233700
A; Map position: 7q11.23-7q11.23
A; Introns: 24/3
A; Note: the list of introns is incomplete
C; Superfamily: neutrophil cytosol factor 1; SH3 homology
C; Keywords: cytosol; neutrophil
F; 163-210/Domain: SH3 homology <SH31>
F; 233-280/Domain: SH3 homology <SH32>

Query Match 99.2%; Score 2018; DB 1; Length 390;
Best Local Similarity 98.7%; Pred. NO. 1.1e-145;
Matches 385; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy	1	MGDTFRHIALLGFEKRFVPSQHYVYMFVWKWDJSEKVVYVRRFTIYIEFHKTLKEMFPI	60
Db	1	MGDTFRHIALLGFEKRFVPSQHYVYMFVWKWDJSEKVVYVRRFTIYIEFHKTLKEMFPI	60
Qy	61	EAGAIINPENIIIPHLPAKWFQDQRAAENRQGTLTIEYCSLMSLPTKISRCPHLLDFVKV	120
Db	61	EAGAIINPENIIIPHLPAKWFQDQRAAENRQGTLTIEYCSLMSLPTKISRCPHLLDFVKV	120
Qy	121	RPDDLPTDQTKKPEYIMPKDGGKSTADITGPILQIOTYRAIANYEKTSGSEMAJSTG	180
Db	121	RPDDLPTDQTKKPEYIMPKDGGKSTADITGPILQIOTYRAIADYEKTSGSEMAJSTG	180
Qy	181	DVVEVVEKSESGWFMCKAKRGWIPASFLPLDSDPETEDPEPNYAGPYVAIKAYTAV	240
Db	181	DVVEVVEKSESGWFMCKAKRGWIPASFLPLDSDPETEDPEPNYAGPYVAIKAYTAV	240
Qy	241	EGDEVSLLEGEAEVYTHKLLDGGWV --- KDDVTGYFPSMWLQKSGQDVSAQRIKRGAPP	296
Db	241	EGDEVSLLEGEAEVYTHKLLDGGWVTRKDDVTGYFPSMYLQKSGQDVSAQRIKRGAPP	300
Qy	297	RRSSIRNAHSIHORSKRRLSODAYRNNSVRFYLOQRROARPGQSPGSPLEEROTORSK	356
Db	301	RRSSIRNAHSIHORSKRRLSODAYRNNSVRFYLOQRROARPGQSPGSPLEEROTORSK	360
Qy	357	POPAPVPPPSADLILNRCSESTKRRKLASAV	386
Db	361	POPAPVPPPSADLILNRCSESTKRRKLASAV	390

RESULT 2
I54525
leukemia-related protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 16-Jul-1999
C:Accession: I54525
R:Jackson, S.H.; Malech, H.L.; Korak, C.A.; Lomax, K.J.; Gallin, J.I.; Holland, S.M.
Immunogenetics 39, 272-275, 1994
A:Title: Cloning and functional expression of the mouse homologue of p47phox.
A:Reference number: I54525; MUID:94164697
A:Accession: I54525
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-388 <RES>
A:Cross-references: GB:I11455; NID:g309422; PIDN:AAA50469.1; PID:g557868
C:Superfamily: neutrophil cytosol factor 1; SH3 homology
F:161-208/Domain: SH3 homology <SH31>

Query Match 81.9%; Score 1656; DB 2; Length 388;
Best Local Similarity 80.2%; Pred. No. 5.7e-119;
Matches 312; Conservative 35; Mismatches 36; Indels 6; Gaps 3;

QY 3 DTFI RIALIGFEKFRVP SQHYVVMFLVKWQDLSEKVVYRRFTEIYEFHKT LKEMFPTEA 62

Dbb

1 DFTFRIALLGFEKRFIPISQHVYVMFLVKWQDLSKVYVRKETETVEFIHKMKEMPPIEA 60
63 GAINPENRIIPLHPAPKWFDGQRAENRQGLTEYCSTIMSLPTIKISRCPHLDDFFKVRP 122
 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 GEIHTENRV IPHLPAPRWFDGORAARERQGLTTEYFNGLMGLPVKISRCPHLDDFFKVRP 120
123 DDKLPTDNQTKPKETYLMMPDKGSTANDITGPIILQTYIRAITANYEKTSGSEMAJLSTGDV 182
 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 DDLLPTDSQAQKPETYLVLPDKGNNAVDITGPPIILOTHRAIDAYEKSSGTGMTATGDV 180
183 VEVYEKSESGWWFCOMKAKRGWIPASFLEPLELSDSPDETPEDPEPNYAGEPYVAIKAYTAVEG 242
 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 VDVEYKESGGWFVCOMKTRGVWPASYILEPLDSPDAEDPDNYAGEPYVIKAYAABEE 240
243 DEVSLLEGEAEVEVIHKKLDGW----KDDVTGYFPFSMYLQKGODVSQAQRQIK-RGAPPR 297
 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
241 DEMSLSEGEATEVIHKKLDGWVVWRKGDITGYFPFSMWYLQKAGEEITQAQRQIRGRGAPPR 300
298 RSSRTNHAHSIHORSKRRLSODAYRNRSVRFLOORRRARPQPSPGPSPLEEERQRTORSKP 357
 | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
301 RSTLRNAQSIIHQSRKRRLSQDYTRYNRNSVRFLOORRRARPRAASTDG-TKDNPSTPRVAP 359
358 QPAYVPPRPSSADLIILNRCSESTKRKLASAV 386
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
360 QPAYVPPRPSSDLILHRCTESTKRKLTSAV 388
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
T00056
hypothetical protein KIAA0418 - human
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C:Accession: T00056
R:Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka,
submitted to the EMBL Data Library, October 1997
A:Description: Prediction of the coding sequences of unidentified human genes. VIII.
A:Reference number: Z14080
A:Accession: T00056
A>Status: preliminary; translated from GB/EMBL/DDBB
A:Molecule type: mRNA
A:Residues: 1-940 <ISH>
A:Cross-references: EMBL:AB007878; NID:d1179754; PIDN:BAA24848.1; PID:d1025770
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0418

Query Match 13.5%; Score 275; DB 2; Length 940;
Best Local Similarity 26.9%; Pred. No. 8e-13;
Matches 70; Conservative 46; Mismatches 86; Indels 58; Gaps 7;

Qy 156 IILQYRAIANYEKTSGSEMA^{LSTGDVVEVVEKSES}GWFQMKAKRGWIPASFTLEPLDS 215
:^{|||}:[|]:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:^{|||}:

Dp 1 MILEOYVVVSNYKKOENSELSLOAGEVDVTEKNESGWFVS^{TSEEQGWVPATYLEAONG} 60

Qy 216 PDTEDEPNYAG--EPYVAIKAYAVEGDEVSLLEGEAVEVTHKLLDGW-----KDD 266

Dh 61 MPBDSQIAMSCKTCSEKQVVMVDVDMSCQKDETCSEKQVIVAVENVTBQNI EQLMIVIBVICKP. 110

[illegible]

Qy 298 -----RSSIRNAHSIHQRSKRRLSQDAYERNRSVFLQRRQRARPGPSPG 343
 : | : | : |:| : | ||:| : | ||:| : | ::

QY 344 SPLEEROTQRSKPQPAVPP 363

```

DD      230 FF KRESSSGEQUFAFFFF 234

```

RESULT 4

T09194
adaptor protein intersectin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09194
R:Yamabhai, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castagnoli, L.; Cesareni
J. Biol. Chem. 273, 31401-31407, 1998
A:Title: Intersectin, a novel adaptor protein with two eps15 homology and five src homol
A:Reference number: Z16605; MUID:99030416
A:Accession: T09194
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1270 <YAM>
A:Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AAC73068.1; PID:g2642625
A:Experimental source: cell type oocyte
C:Function:
A:Description: involved in endocytosis
C:Keywords: endocytosis

Query Match 8.8%; Score 179.5; DB 2; Length 1270;
Best Local Similarity 21.1%; Pred. No. 2.2e-05;
Matches 93; Conservative 66; Mismatches 135; Indels 147; Gaps 23;

Qy 38 KVVYRRTTEIYEFHKTLEKMPPIEAGINPENRIIPHPAPKWFQD----- 83
Db 733 KVVYR--ALYPFDARSHDEITIEPGDIIMVDE--SQTGEPCWLGGLKGTGWFPANYA 788

Qy 84 ORAAENROGTLTEVCSTLMSLPT-KISRCP-----HLLDFKVRPDDLKLP 128
Db 789 ERMPSEFPSTTKPAAEETAKPTVHVAPSPVAPAAFTNTSTNSNWADFSTWP----- 842

Qy 129 TDNOKKPE-----TYLMPKDGKS-----TATDITG---PIILQ-----T 160
Db 843 -TNNTDKVESDNWTWAAQPSLTVPSSAGHQRAFTPATVTVGSSPSVLCQGEKVEGLQ 901

Qy 161 YRAIANEKTSGSEMASTGSDVVEVEKESGWFQCMKARKGIPASFLEPLDSP----- 216
Db 902 AQALYPWRAKKDNHNFKNNDVITVLEQDM-WWFEVQGGKGFPSKYVVLISGLPKRS 960

Qy 217 --DET--EDP-----EPNYAGPVAIKAYTAVGDEVSLLEGAVEVIHKL 260
Db 961 TSIDSTSESASLKRVSPPAFKPAIOGEEYISMVYTESNQGLITFOQGLIIVVKKDG 1020

Qy 261 DGWKDDV---TGYPSPMYLQ-----KSGQ-----DVSO-----AQRQIKRG 293
Db 1021 DWVTGTVEKTVGVPSPNIVYRPKDSAGSGGKTSGLKKPEIAQVIASVYATAPQLPL- 1079

Qy 294 APPRRSSIRNAH-----SIHRSRKRRLSDAYRRNSVRFLOQRRRQARPQSPGSP 346
Db 1080 APGQILIRKKNPGGWMBGELQARGKKR-QIGWFPANYVKLL-----SPGT-- 1124

Qy 347 EERQRTQSRKQPAVPPRPSA 367
Db 1125 -----NKSTPEPPKPTS 1137

RESULT 5
S39768
neutrophil cytosol factor p40 - human
C:Species: Homo sapiens (man)
C>Date: 19-May-1994 #sequence_revision 17-Nov-1995 #text_change 07-May-1999
C:Accession: S39768; S39769
R:Wientjes, F.B.; Hsuan, J.J.; Totty, N.F.; Segal, A.W.
Biochem. J. 296, 557-561, 1993
A:Title: p40(phox), a third cytosolic component of the activation complex of the NADPH o
A:Reference number: S39768; MUID:94107216
A:Accession: S39768
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-339 <WIE>
A:Accession: S39769
A:Molecule type: protein

A:Residues: 52-92;115-125;131-145;146,176-186;192-213;228-234;241-263;297-308;317-322
C:Superfamily: SH3 homology
C:Keywords: cytosol
F:177-224/Domain: SH3 homology <SH3>

Query Match 8.3%; Score 169.5; DB 2; Length 339;
Best Local Similarity 24.4%; Pred. No. 2e-05;
Matches 76; Conservative 48; Mismatches 116; Indels 71; Gaps 12;

Qy 15 EKRFVPSOHVYVMELVKWQDLSEKVVYRFTTEIYEFHKTLEKMPPIEAGINPENR--- 70
Db 31 EKRGFTS-HFVFVIEVTKGSKYLIYRRYRQFHALQSKLEERF-----GPDSKSSAL 82

Qy 71 --IIPHPAPKWFQD--ORAAENROGTLTEVCSTLMSLPTKISRCPHLLDFKVRP-DDL 125
Db 83 ACTLPTLPA-KVYGVKQIEIAEMRIPALNAYMKSLSLPVMVLMDEDVRIFFYQSPDSE 141

Qy 126 KLPTDNQTKPKETYLMPKDGKSTATDITGPILQYIRAIANYEKTSGSEMALS--TGDVV 183
Db 142 QVPAIRRLRPRT-----RKVKSVPQGSVDMAAPRAEALFDFTGNSKLELNFKAGDVI 197

Qy 184 EVVEKESGWFQCMKARKGIPASFLEPLDSPDTEDP----- 222
Db 198 FLLSRINKDWLEGTVRGATGIRPLSFVKILKDPPEEDPTNMLRCYVYEDTISTIKDIAV 257

Qy 223 EPNVAGEPY-----VAIKAYTAVGDEVSLLEGAVEVI-----H 257
Db 258 EEDLSSTPLKDLLELTRREFQREDIALN-YDAEGDLVRLLSDEDVALMVRQAGLPSQ 316

Qy 258 KLLDGWKDDVT 268
Db 317 KRLFPWKLHIT 327

RESULT 6
T13151
adaptor protein CMS - human
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T13151; T08754
R:Kirsch, K.H.; Georgescu, M.M.; Ishimaru, S.; Hanafusa, H.
Proc. Natl. Acad. Sci. U.S.A. 96, 6211-6216, 1999
A:Title: CMS: An adaptor molecule involved in cytoskeletal rearrangements.
A:Reference number: Z17608; MUID:99272673
A:Accession: T13151
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-639 <KIR>
A:Cross-references: EMBL:AF146277; NID:g4960046; PID:g4960047; PIDN:AAD34595.1
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08754
A:Molecule type: mRNA
A:Residues: 548-639 <WAM>
A:Cross-references: EMBL:AL050105
A:Experimental source: adult uterus; clone DKFZp586H0519
C:Genetics:
A:Note: DKFZp586H0519.1
C:Complex: homodimer
C:Function:
A:Description: probably functions as a scaffolding molecule with a specialized role 1
C:Keywords: coiled coil; homodimer

Query Match 8.1%; Score 164.5; DB 2; Length 639;
Best Local Similarity 22.9%; Pred. No. 0.00011;
Matches 86; Conservative 50; Mismatches 127; Indels 113; Gaps 17;

Qy 45 TEIYEFHKTLEKMPPIEAGINPENRIIPHPAPKWFQDQRAAENRQGTPLTEYCSLMSL 104
Db 168 TDDGETHEAQDDSETVLAGPTSP----IPSL-----GNVSETASGSVTQ 207

Query Match 7.4%; Score 150.5; DB 2; Length 665;
Best Local Similarity 24.3%; Pred. No. 0.0014;
Matches 43; Conservative 34; Mismatches 51; Indels 49; Gaps 7;

Qy 166 NYEKTSGSEALSTGDDVVEVESGWWFCOMKAKRGWIPASFLPELDPDPTEDPEPN 225
Db :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
9 DYQAQHDELITISVGELITNIRKDEGGWGQGNGRRGLFPDNFVREI-KKMKKDPLTN 67

Qy 226 YAGEP-----YVAIKAYTAVEGDEVSLLEGSAVE 254
Db | | KAPEKLHEVPVSGNSLLSSETILTRNGERRRRCQAF-SYLQNDDLELKVGDIIE 126

Qy 255 VIHLKLDGWDDV----TGYPSPMYLOK-SGD-----VSQAQRQIKRGAPRRSSIR 302
Db |:|: |:|: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 127 VGVGEVGWEGVLNGTKTFMPFSFIKELSGESDELGSQEQLSK-----SSLR 176

RESULT 10
T42526
hypothetical protein - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42526
R:Yoshiooka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MUID:98162722
A:Accession: T42526
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-290 <YOS>
A:Cross-references: EMBL:D89164; NID:g1749535; PIDN:BAA13826.1; PID:g1749536
A:Experimental source: strain PR745

Query Match 7.2%; Score 147.5; DB 2; Length 290;
Best Local Similarity 20.7%; Pred. No. 0.00074;
Matches 59; Conservative 56; Mismatches 113; Indels 57; Gaps 11;

Qy 5 FIRHIALLGKEKRPVPSQHYVMFLVKQDLSEKVYRRTEIYEF-HKTLKMEPIEAG 63
Db |:|: |:|: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
51 FLRGLLPFLYQKPRSNHLLLHLHFQFLNLHLLLLLYLKLFQFLNHLLPPYLRLFLQ 110

Qy 64 AINPENRIIPHLPAPKWFDGORAENRQGTLEYCYSTMSLPTKRISCPHLLDFPKVRPD 123
Db | | FLKPP--AVPVVPEP-----AGOLNE-----PV 131

Qy 124 DLKLPTNQKKPTYLMPKDGKSTATDTGPTILOYYRAIANYEKTSGSEMAJSTGDVV 183
Db ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 132 VPPLPHDETQEPQV-----GGDVKAITEHTOPTKTPAI-VIYDYSPEENEIELVENEQI 185

Qy 184 EVVEKSESGWFFCO-MKAKRGWIPASFLPELDPDPTED--PEPNYAGEPVVALKA--Y 237
Db |:|: |:|: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 186 QILEFVDGHWMLGNSKGQGLFFPSNVIE-ITGNETANNPPAEPPAQAGPGKSVKRAYDY 244

Qy 238 TAVEGDEVSLLEGSAVEIVHKLDGWKDDV----TGYPSPMYLOK 278
Db | | |:|: |:|: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 245 QAQEDNELSFEDIELIANVCVDPNWWEGETGHGRGLFPSNYVEE 289

RESULT 11
T50995
related to cytoskeleton assembly control protein SLAL [imported] - Neurospora crassa
N:Alternate names: protein B7f18.140
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50995
R:Schulze, U.; Algn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25286
A:Accession: T50995
A>Status: preliminary

Db 1 MGOTFIKHIALLGFEKRFVPSQHYVHV 27

RESULT 15

T17257

hypothetical protein DKFzp586p1422.1 - human

C:Species: Homo sapiens (man)

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17257

R:Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17257

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-816 <KOE>

A:Cross-references: EMBL:AL117472

A:Experimental source: adult uterus; clone DKFzp586p1422

C:Genetics:

A:Note: DKFzp586p1422.1

Query Match .. 6.7%; Score 135.5; DB 2; Length 816;

Best local Similarity 19.3%; Pred. No. 0.026;

Matches 73; Conservative 57; Mismatches 113; Indels 135; Gaps

QY 32 WDLSEKVVYRRF---TEIYFHKTLKEMFPIE--AGAINPENRIIPHLF---APKWF 81

Db 438 WEPPDKKVDTRKYRAEPKSIYEQPKSVLTNEKMSSAISFTPEISSETPGYIYSSNFH 497

QY 82 DGQRAAENRQGTLT-----YCSTL-----MSLPTKISRCP-HLL- 115

Db 498 AVKRESGAPGDLTSLNENRQIYKSVLEGGDIPLQGLSGLKRPPSSASTKDSERHPFIP 557

QY 116 -----DFFKVRPDD-----LKLPTDQTKK----- 135

Db 558 ADYLESTEFIRRHDDKEMRPARAFKQATKELPLQKGDIVYIKQIDQNWYGEH 617

QY 136 -----PETY---LMPKDGKSTATDITGPILIQTYRAIA--NYEKTSGSEMASTGDWVE 184

Db 618 HGRVGIFPRTYIELLPAPAEKAQPKLTPQVLEYGEALAKENFNGDQTQVENSFRKERIT 677

QY 185 VVKSSEGNWFQCM--KAKRGWIPASFLE-----PLD----- 214

Db 678 LLRQVDENWYEGRIPTGSRQGFITYVDVYIKRPLVKPVDMYMDLPFSSSPSRSATASQ 737

QY 215 -----SPDTEDPENYAGEPVAKAYATVAGEGVSLLGEAEVEVIHKLDGM-- 263

Db 738 OPAQAQRVTPDRSQTSQDLFS---YQADYSIYIPONDDELELRDGDIVDMVEKCDGWFV 794

QY 264 ----KDDVTGYFPFMYLQ 277

Db 795 GTSRRTKQFGTFPGNYVK 812

Search completed: January 11, 2002, 09:08:16

Job time: 142 sec

Search completed: January 11, 2002, 09:08:16
Job time: 142 sec

Query Match	6.7%	Score 136;	DB 4;	Length 50;
Best Local Similarity	92.6%;	Pred. No. 0.00048;		
Matches 25;	Conservative	2;	Mismatches 0;	Gaps 0;
			Indels	

QY 1 MGDTFIRHIALLGFEKRFVPSQHYVYM 27
|||

THIS PAGE BLANK (USPTO)

Printed 10/767341

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2002, 09:05:54 ; Search time 25.33 seconds
(without alignments)
1128.792 Million cell updates/sec

Title: US-09-820-005-2
Perfect score: 2035
Sequence: 1 MGDTRHALLGFEKRFVP.....ADILNRCSESTKRKLASAV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_1101.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2006	98.6	390	16	AA83825
2	691	34.0	128	22	AA64031
3	688	33.8	128	22	AA64033
4	593	29.1	141	21	AA833681
5	305	15.0	64	22	AA64032
6	295	14.5	64	22	AA64034
7	261	12.8	968	22	AAU14174
8	188	9.2	509	17	AAW05399
9	187	9.2	438	21	AA840983
10	183	9.0	1683	21	AA711160
11	177.5	8.7	1197	21	AA57445

12	177.5	8.7	1658	21	AA57450	Mouse Ese2L protei
13	176.5	8.7	248	17	AAW05396	Human SH3P18 prote
14	172	8.5	1214	21	AA57444	Mouse Esel protein
15	168.5	8.3	1715	21	AA57449	Mouse EsellL protei
16	167	8.2	1215	20	AA572156	Human SH3D1A prote
17	166	8.2	1220	20	AA572155	Human SH3D1A prote
18	157	7.7	462	17	AAW05395	Human SH3P17 prote
19	157	7.7	641	20	AA572158	Human SH3D1A prote
20	157	7.7	1144	20	AA572154	Human SH3D1A prote
21	156.5	7.7	709	21	AA572154	Human SH3D1A prote
22	150.5	7.4	665	21	AA572154	Human SH3D1A prote
23	150	7.4	665	21	AA572154	Human SH3D1A prote
24	147.5	7.2	324	18	AAW26497	CD2-associated int
25	147.5	7.2	324	18	AAW25117	CD2-associated int
26	147.5	7.2	324	22	AAW80421	CD2-associated int
27	147.5	7.2	324	22	AAW66392	Human CD2 associat
28	144.5	7.1	635	19	AAW48898	Candida albicans C
29	144	7.1	464	18	AAW26496	CD2-associated int
30	144	7.1	464	18	AAW25116	CD2-associated int
31	144	7.1	464	20	AAW80420	CD2-associated int
32	144	7.1	464	22	AAW66391	Human CD2 associat
33	141	6.9	553	18	AAW26495	CD2-associated int
34	141	6.9	553	18	AAW25115	CD2-associated int
35	141	6.9	553	20	AAW80419	CD2-associated int
36	141	6.9	553	22	AAW66390	Human CD2 associat
37	141	6.9	659	19	AAW37724	CD2-associated int
38	137.5	6.8	397	20	AAW28292	Amino acid sequenc
39	137.5	6.8	416	20	AAW28291	Amino acid sequenc
40	137.5	6.8	416	21	AAW68782	Amino acid sequenc
41	134.5	6.6	1035	22	AAW43519	Human polypeptide
42	133.5	6.6	370	21	AAW99428	Human PRL1431 (UNQ
43	133.5	6.6	370	22	AAW66177	Protein of the inv
44	133.5	6.6	1200	21	AAW19313	Amino acid sequenc
45	132.5	6.5	415	19	AAW1595	Murine tyrosine ph

ALIGNMENTS

RESULT	1	
AA83825		
ID	AA83825	standard; Protein; 390 AA.
XX		
AC	AA83825;	
XX		
DT	15-FEB-1996	(first entry)
XX		
DE	p47(phox)	protein.
XX		
KW	p47(phox);	p67(phox); NADPH oxidase complex; proline-rich region; PTPase;
KW	SH3 domain;	inhibition; dynamin; cytochrome b245; reperfusion injury;
KW	septic shock;	arthritis; asthma; vinculin; inflammatory bowel disease;
KW	adult respiratory	distress syndrome; ischaemic heart disease.
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Misc-difference 22	/label= OTHER
FT		/note= "designated O in specification"
FT		
FT	Misc-difference 197	/label= OTHER
FT		/note= "designated O in specification"
FT		360..373
FT	Region	/note= "proline-rich region used to design inhib"
XX		
PN	GB2285047-A.	
XX		
PD	28-JUN-1995.	
XX		
PF	07-DEC-1994;	94GB-0024674.
XX		
PR	07-JAN-1994;	94GB-0000248.


```
PR 21-DEC-1993; 93GB-0026083.
XX (LUDW-) LUDWIG INST CANCER RES.
PA (YAMA ) YAMANOUCHI UK LTD.
XX
XX Finan PM, Gout IT, Kellie S, Shimizu Y, Waterfield MD;
PI WPI; 1995-217703/29.
XX
XX New polypeptide(s) with antiinflammatory action - inhibit NADPH
PT oxidase system.
XX
XX Example 1; Fig 1; 17pp; English.
XX
XX The protein p47(phox) (AAR83825) interacts with protein p67(phox) in the
CC NADPH oxidase complex via a C-terminal proline-rich region of p47(phox)
CC binding to at least one C-terminal SH3 domain of p67(phox). The
CC peptides AAR83814-24 were derived from the proline-rich region of
CC p47(phox) and show inhibitory activity towards the binding of p47(phox)
CC to p67(phox) as compared to the proline-rich region of a range of other
CC proteins (AAR83826-30). The inhibitory peptides can be used for the
CC treatment of chronic or acute inflammatory diseases e.g. septic shock,
CC arthritis, asthma, adult respiratory distress syndrome, ischaemic heart
CC disease, reperfusion injury or inflammatory bowel disease.
XX
XX Sequence 390 AA;
SQ
Query Match 98.6%; Score 2006; DB 16; Length 390;
Best Local Similarity 98.2%; Pred. No. 8.5e-187;
Matches 383; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
Qy 1 MGDTFIRHALLGTEKRPVPSQHYVYMLVKQDLSEKVVYRRTEIYEFHKTLEMPPI 60
Db 1 mgdtfirhiallgfkrfvpsqhyvymflvkwdlsekvvrrrfefyefhktlkempfi 60
Qy 61 EAGAINPENRIIPLHLPAPKWFQDGRAAENRQGTLETCSTLMSLPTKISRCPHLLDFFKV 120
Db 61 eagainpenriiiphlpapkwfdgqraaenrggtlteycstlmslptksrphllldffkv 120
Qy 121 RPDDLKLP 128
Db 121 rpddlklp 128
Qy 121 RPDDLKLP 128
Db 121 rpddlklp 128
Query Match 34.0%; Score 691; DB 22; Length 128;
Best Local Similarity 100.0%; Pred. No. 2e-59;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGDTFIRHALLGTEKRPVPSQHYVYMLVKQDLSEKVVYRRTEIYEFHKTLEMPPI 60
Db 1 mgdtfirhiallgfkrfvpsqhyvymflvkwdlsekvvrrrfefyefhktlkempfi 60
Qy 61 EAGAINPENRIIPLHLPAPKWFQDGRAAENRQGTLETCSTLMSLPTKISRCPHLLDFFKV 120
Db 61 eagainpenriiiphlpapkwfdgqraaenrggtlteycstlmslptksrphllldffkv 120
Qy 121 RPDDLKLP 128
Db 121 rpddlklp 128
RESULT 3
AAG64033
ID AAG64033 standard; protein; 128 AA.
XX
XX AAG64033;
AC AAG64033;
DT 07-SEP-2001 (first entry)
XX
XX Human PX domain-related protein.
XX
XX Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH;
KW NADPH oxidase; superoxide production; inflammatory disease.
XX
```

```
OS Homo sapiens.
XX WO200142453-A1.
XX 14-JUN-2001.
XX
XX 01-DEC-2000; 2000WO-JP08501.
XX
XX 06-DEC-1999; 99JP-0346193.
XX
XX (BIOM-) BIOMOLECULAR ENG RES INST.
XX
XX Kohda D, Hiroaki H, Sumimoto H;
XX WPI; 2001-381679/40.
XX
XX Regulation of function of a protein containing a PX domain for
XX controlling an inflammatory response to disease or trauma, comprises
XX using nuclear magnetic resonance (NMR) chemical shift and structural
XX coordinate data -
XX
XX Example 3; Page 186; 195pp; Japanese.
XX
XX The present sequence is provided in a specification relating to the use
XX of chemical shift data from nuclear magnetic resonance (NMR) and
XX structural coordinate data of the PX domain of a protein to search for,
XX evaluate, design and identify variant PX domain sequences for
XX controlling the function of proteins containing the PX domain. It also
XX relates to compounds promoting the binding of substances to the PX
XX domain, and compounds inhibiting the binding of substances to the PX
XX domain. The invention can be used in the identification of substances
XX that can be used in the regulation of reduced nicotinamide adenine
XX dinucleotide phosphate (NADPH) oxidase activity, which is involved in
XX superoxide production in inflammatory response in disease or trauma.
XX
XX Sequence 128 AA;

Query Match 33.8%; Score 688; DB 22; Length 128;
Best Local Similarity 99.2%; Pred. No. 4e-59;
Matches 127; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGDFTIRHIALGFEKRVPSQHYVYMFVLKQDLSEKVVYRRFTIYEFHKLKEMFPI 60
Db 1 mgdtfirhiallgfekrvpsqhyvymflvkqdlsekvvyrrfteiyefhklkemfpi 60

Qy 61 EAGAINPENRIIPLPAPKWFDCQRAAENRQGLTEYCSLTMSLPTKISRCPHLLDFKVP 120
Db 61 esgainpenriiiphlpapkwfdcqraaenrqglteycstlmslptkiskrcphllldfkvp 120

Qy 121 RPDDLKLP 128
Db 121 rpddlklp 128

RESULT 4
AAB53681
ID AAB53681 standard; Protein; 141 AA.
XX
XX AAB53681;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1221.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX identification; cytostatic; cardioactive; neuroprotective; vulnery;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;
XX infectious disease; cardiovascular disorder.
XX
```

```
OS Homo sapiens.
XX WO2000055351-A1.
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX N-PSDB; AAC98438.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer -
XX
XX Claim 11; Page 1810-1811; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnery, nephrotropic, antiinfective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies to the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins
XX may also be used to prevent diseases such as neural disorders, immune
XX system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, wounds, renal disorders, infectious
XX diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX AAB54007 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 141 AA;

Query Match 29.1%; Score 593; DB 21; Length 141;
Best Local Similarity 70.7%; Pred. No. 8.3e-50;
Matches 118; Conservative 7; Mismatches 16; Indels 26; Gaps 2;

Qy 3 DTFIRHIALGFEKRVPSQHYVYMFVLKQDLSEKVVYRRFTIYEFHKLKEMFPIEA 62
Db 1 dtfirhiallgfekrvpsqhyv-----hvpgemagpvge 35

Qy 63 GAINPENRIIPLPAPKWFDCQRAAENRQGLTEYCSLTMSLPTKISRCPHLLDFKVP 122
Db 36 ggalpalhrdl-rvpspkwfdgqraaenhqgtlteycstlmslptkiskrcphllldfkvp 94

Qy 123 DDLKLPDNTQTKKPTYLMPKDGKSTATDITGPILQTYRAIANYEK 169
Db 95 ddllkpldntqtkkptylmpkdgkstatditegpilqtyraianyek 141

RESULT 5
AAG64032
ID AAG64032 standard; protein; 64 AA.
XX
XX AAG64032;
XX
XX 07-SEP-2001 (first entry)
XX
XX Human SH3 domain.
XX
XX Human; antiinflammatory; protein coordinate data; chemical shift data;
XX nuclear magnetic resonance; NMR; structural coordinate data;
XX PX domain; reduced nicotinamide adenine dinucleotide phosphate; NADPH;
XX
```

KW NADPH oxidase; superoxide production; inflammatory disease; SH3 domain.
XX
OS Homo sapiens.
XX
PN WO200142453-A1.
XX
PD 14-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-JP08501.
XX
PR 06-DEC-1999; 99JP-0346193.
XX
PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX
PI Kohda D, Hiroaki H, Sumimoto H;
XX
DR WPI; 2001-381679/40.
XX
PT Regulation of function of a protein containing a PX domain for
PT controlling an inflammatory response to disease or trauma, comprises
PT using nuclear magnetic resonance (NMR) chemical shift and structural
PT coordinate data -
XX
PS Example 2; Page 183; 195pp; Japanese.
XX
CC The present sequence is provided in a specification relating to the use
CC of chemical shift data from nuclear magnetic resonance (NMR) and
CC structural coordinate data of the PX domain of a protein to search for,
CC evaluate, design and identify variant PX domain sequences for
CC controlling the function of proteins containing the PX domain. It also
CC relates to compounds promoting the binding of substances to the PX
CC domain, and compounds inhibiting the binding of substances to the PX
CC domain. The invention can be used in the identification of substances
CC that can be used in the regulation of reduced nicotinamide adenine
CC dinucleotide phosphate (NADPH) oxidase activity, which is involved in
CC superoxide production in inflammatory response in disease or trauma.
XX
SQ Sequence 64 AA;

Query Match 15.0%; Score 305; DB 22; Length 64;
Best Local Similarity 93.8%; Pred. NO. 2.8e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 223 EPNYAGEPYVAIKAYTAVEGDEVSLLEGEAVFIHKLIDG-W---KDDVTGYFSPMYLQK 278
Db 1 epyagepyvaikaytavegdevslllegeavevindhklidgwwvirkddvtgyfpsmylqk 60

QY 279 SQGD 282
Db 61 sgqd 64

RESULT 6
AAG64034
ID AAG64034 standard; Protein; 64 AA.
XX
AC AAG64034;
XX
DT 07-SEP-2001 (first entry)
XX
DE Human SH3 domain-related protein.
XX
KW Human; antiinflammatory; protein coordinate data; chemical shift data;
KW nuclear magnetic resonance; NMR; structural coordinate data; PX domain;
KW reduced nicotinamide adenine dinucleotide phosphate; NADPH; SH3 domain;
KW NADPH oxidase; superoxide production; inflammatory disease.
XX
OS Homo sapiens.
XX
PN WO200142453-A1.
XX
PD 14-JUN-2001.

XX
PF 01-DEC-2000; 2000WO-JP08501.
XX
PR 06-DEC-1999; 99JP-0346193.
XX
PA (BIOM-) BIOMOLECULAR ENG RES INST.
XX
PI Kohda D, Hiroaki H, Sumimoto H;
XX
DR WPI; 2001-381679/40.
XX
PT Regulation of function of a protein containing a PX domain for
PT controlling an inflammatory response to disease or trauma, comprises
PT using nuclear magnetic resonance (NMR) chemical shift and structural
PT coordinate data -
XX
PS Example 4; Page 188; 195pp; Japanese.
XX
CC The present sequence is provided in a specification relating to the use
CC of chemical shift data from nuclear magnetic resonance (NMR) and
CC structural coordinate data of the PX domain of a protein to search for,
CC evaluate, design and identify variant PX domain sequences for
CC controlling the function of proteins containing the PX domain. It also
CC relates to compounds promoting the binding of substances to the PX
CC domain, and compounds inhibiting the binding of substances to the PX
CC domain. The invention can be used in the identification of substances
CC that can be used in the regulation of reduced nicotinamide adenine
CC dinucleotide phosphate (NADPH) oxidase activity, which is involved in
CC superoxide production in inflammatory response in disease or trauma.
XX
SQ Sequence 64 AA;

Query Match 14.5%; Score 295; DB 22; Length 64;
Best Local Similarity 93.8%; Pred. NO. 2.7e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 4; Gaps 2;

QY 223 EPNYAGEPYVAIKAYTAVEGDEVSLLEGEAVFIHKLIDG-W---KDDVTGYFSPMYLQK 278
Db 1 epyagepyvaikaytavegdevslllegeavevindhklidgrrwvirkddvtgyfpsmylqk 60

QY 279 SQGD 282
Db 61 sgqd 64

RESULT 7
AAU14174
ID AAU14174 standard; Protein; 968 AA.
XX
AC AAU14174;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #45.
XX
KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-451939/48.
XX N-PSDB: AAS22479.
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
XX Example 4; Page 551-552; 894pp; English.
XX The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/ elicit an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence represents a protein of the invention.
XX Sequence 968 AA;

Query Match 12.8%; Score 261; DB 22; Length 968;
Best Local Similarity 24.3%; Pred. No. 3.5e-16;
Matches 70; Conservative 46; Mismatches 86; Indels 86; Gaps 7;
QY 156 ILIOTYRIANYKTSSEMALSTGDDVVEVVEKESGHWFCQAKRGWIPASLEPLDLS 215
Db 1 mllqgyvvvnykqnselslqgevvvdknesgwwfvstseegwvpatyleaqq 60
QY 216 PDETEDEPVPYAG-----EPYVAIKAYTAVEGDEV 245
Db 61 trdssdintstgevsrkrkahlrldrrwtlqgmvrqrhsrekyvtvqytsqskdei 120
QY 246 SLLGGEAVEVTHKLLDGW-----KDDVTGYFFSMYLOKSGQDVSAQROI----- 290
Db 121 gfekgvtvevirknlegwvrylygke---gwapasyllkakddlptrkknlagpveilig 177
QY 291 -----KRGAPR-----RSIRNAHSIHQSRKRL 315
Db 178 nimeisnllnkasgdketpaaegheapiakeislplilcnsdgsavgvdpdrtvsrl 237
QY 316 SQDAYRNSVRFLOQRQARPGPQSGPSPLEERQTORSPQPAVPP 363
Db 238 aggs--pavariapraqisnrltrppp-rresslgfqlpkppepp 282

RESULT 8
AAW05399
ID AAW05399 standard; Protein; 509 AA.
XX
AC AAW05399;
XX

DT 19-FEB-1998 (first entry)
XX Human clone 65 protein.
DE
XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW cellular signalling element; cellular structural element; malignancy;
KW protein identification; functional domain; protein screening;
KW cellular signal transduction process.
XX Homo sapiens.
XX OS
XX WO9631625-A1.
XX 10-OCT-1996.
XX PD
XX 04-APR-1996; 96WO-US04454.
XX PF
XX 03-APR-1996; 96US-0630915.
XX PR 07-APR-1995; 95US-0417872.
XX PA (CYTO-) CYTOGEN CORP.
XX PA (UYNC-) UNIV NORTH CAROLINA.
XX PI Fowlkes DM, Hoffman N, Kay BK, McConnell SJ, Sparks AB;
XX WPI: 1996-465045/46.
XX DR N-PSDB: AAT39799.
XX XX Identifying polypeptide(s) having specific functional domain (esp.
PT SH3 domain) - comprises detecting selective binding to recognition
PT unit, regardless of sequence homology
XX Claim 54; Fig 59; 174pp; English.
XX AAW05386-W05403 represent novel human and mouse Src-homology region 3
CC (SH3) domain containing proteins that can be used in the method of the
CC invention. SH3 domain containing proteins play a role in signalling and
CC structural elements of cells. The method of the invention is for
CC identifying polypeptides containing functional domains of interest
CC (especially SH3 domains). The method comprises contacting a multivalent
CC recognition unit (RU) complex with a number of peptides and identifying
CC polypeptides having a selective binding affinity for the RU complex. The
CC method is based on functional similarities and does not rely on sequence
CC similarities. Prior methods only gave limited success for identifying
CC proteins which contain an SH3 domain due to the minimal sequence
CC homology among known SH3 proteins. It has been found that small peptide
CC RUS in multivalent form have reduced specificity for a given functional
CC domain compared to monomer RUS. Multivalent RU complexes are particularly
CC suited to screening for polypeptides containing functional domains that
CC are similar to, but not identical in sequence to, the original target
CC functional domain. The new method enables proteins having a common
CC function to be identified. Identification of novel SH3 proteins will be
CC useful for a better understanding of cell growth, malignancy, signal
CC transduction processes, etc. New candidate drugs can be identified, and
CC their specificities (e.g. pharmacological activities) can be assessed
CC using the method of the invention.
XX Sequence 509 AA;

Query Match 9.2%; Score 188; DB 17; Length 509;
Best Local Similarity 22.1%; Pred. No. 1.7e-09;
Matches 60; Conservative 58; Mismatches 102; Indels 52; Gaps 9;
QY 64 AINPENRIIPLHPAPKWDGORAENROGTLTEYCSLMSLPTKISRCPHLLDFKVRPD 123
Db 141 avspkallp-----ptvslsatstssepslssnqpsvtdyqnvafs 182
QY 124 DLKPLTDNOTKKPETYLMPKDGKSTATDITGPI-----ILQTYR--AIANYEKTSGSMA 176
Db 183 nltvntswgkksaft-----rtvsgvsvpihgqgvgvnlkaqalcswtakdnhln 235
QY 177 LSTGDVVEVVEKESGWMWFCQAKRGWIPASFLPEPLDSPD-ETEDPBNYA----- 227

Db 236 fskhdilvleqqen-wwfgevhrgrwfkpsvkiilpgsevkrreepealyaavnnkpts 294
Qy 228 -----GEPIYAIKATAYVGEDEVSLLEGEAVEVIAHKLLDGMK---DDVTGYFPSPMYLQKS 279
Db 295 aaysvgeeyialypssvepgdlfttegeellvtqkdgewtgsldrgsifpsnyvkpk 354
Qy 280 QDVVSQAORQIKRGAPPRSSIRNAHSIHORS 311
Db 355 dquesfgsas--ksgasnkpkelaqvtasayvas 384

RESULT 9
AAB40983
ID AAB40983 standard; Protein: 438 AA.
XX AAB40983;
AC
DT 08-FEB-2001 (first entry)
XX Human ORFX, ORF747 polypeptide sequence SEQ ID NO:1494.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; erythematosis; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosis; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
PN
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
PI
DR WPI; 2000-602362/57.
DR N-PSDB; AAC75192.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 1240-1241; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosis, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 438 AA;

Query Match 9.2%; Score 187; DB 21; Length 438;
Best Local Similarity 23.6%; Pred. No. 1.7e-09;
Matches 81; Conservative 47; Mismatches 143; Indels 72; Gaps 15;

Qy 52 KTLKEMFPIEAGAINPENRIIPLHLPAPKWFQDQRAAENROCTITEYGSTLMSLPTKISRC 111
Db 16 knsgeplppkpgpspsn-----pgaldldgvrqnavgrekellssqrd----- 61
Qy 112 PHLLDFFKVRP---DDLK--LPTDNTQTKKPTYLMPKDGKSTATDITGPILQIYRAIAN 166
Db 62 ----grfegrpvpdgdkqrspkmrqpprrdmtiprglnlppkpppgveeeyytiae 117
Qy 167 YKTSSEMALSTGDVVEVEKSEGWFCQMKAGRWIPASFLPLELSDPDEDPENY 226
Db 118 fttipdgisfqaglkveieknlsgwwyiqidkegwapaftidkykksnas--fpnf 175
Qy 227 -AGBPY----VAIKAYAVEGDEVSLLEGEA---VEVIHKLLDG---WKDDVTGYFPSPMY 275
Db 176 laplphvqtirlgeaaalennttseatgsrplpdaphgvmdsglpwskdwkg--skdv 233
Qy 276 LQKSGQD-----VSQAORQIKRGAPPRSS--IRNAHSIHORSKRRLSQDAYRRNSV 325
Db 234 lrkassdmsasagyeicisdpmeeekpslpprkesliksegellerer----- 282
Qy 326 RFLQORRQAR-----PG---PQSPGSPLEERQTORSKPOP 359
Db 283 ----grteqlrgptkpgpvilpmpakhippardsrpepkp 321

RESULT 10
AAY71160
ID AAY71160 standard; Protein: 1683 AA.
XX
XX AAY71160;
XX
XX 08-SEP-2000 (first entry)
XX Rat phosphodiesterase interacting protein, M14.
XX
XX Rat; phosphodiesterase interacting protein; M14; PDE; cAMP-PDE;
KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;
KW antiinflammatory; antipsoriatic; dermatological; antibacterial; shock;
KW analgesic; immunosuppressive; antiulcer; vasotropic; antiarthritic;
KW antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;
KW antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;
KW atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
KW eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
KW reperfusion injury; atopic dermatitis; diabetes insipidus;
KW conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
KW arterial restenosis; ankylosing spondylitis; transplant rejection;
XX graft versus host disease.
OS Rattus sp.
XX WO200027861-A1.
PN
XX 18-MAY-2000.
PD
XX

```
PF 12-NOV-1999; 99WO-US26860.
XX
PR 12-NOV-1998; 98US-0108255.
XX
PA (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI Conti M, Pahlke G;
XX
DR WPI; 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
PT polypeptide, useful for diagnosis and treatment of asthma, cystic
PT fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
PS Disclosure: Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
CC M14 from rat. The protein modulates the functions and properties of PDEs,
CC specifically cAMP-PDEs, and also targets them to specific subcellular
CC compartments. The present sequence
CC can be used in the diagnosis and treatment of disease conditions
CC associated with PDE activity. The diseases include asthma, cystic
CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
CC granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic
CC conjunctivitis, vernal conjunctivitis, arterial restenosis,
CC atherosclerosis, inflammatory diseases associated with irritation and
CC pain, rheumatoid arthritis, ankylosing spondylitis, transplant
CC rejection and graft versus host disease, disease conditions associated
CC with hypersecretion of gastric acid, and disease conditions in which
CC cytokines are mediators.
XX
SQ Sequence 1683 AA;

Query Match 9.08; Score 183; DB 21; Length 1683;
Best Local Similarity 28.08; Pred. No. 3.2e-08;
Matches 47; Conservative 37; Mismatches 62; Indels 22; Gaps 6;

Qy 162 RAIANEYKTSSEALSTGDDVVEVEKSEGWFCQKAKRGWIPASFLPLELSDPD-ETE 220
Db qalcswtakkenhlnfskhdvltvledqen--wfigevhgrgwfpskypvkiipgsevkrg 948

Qy 221 DPEPNYA-----GEPYVAIKAYTAVEGDEVSLLGEAVEVHKLDDGWKDDV 267
Db epealyaavnkktstaypvgeeyalyssvpegdltftegeellvtqkdgewtgtsi 1008

Qy 268 ---TGYPFPMYLO-KSGODVYSAQRQIKRGAPRRSSIRNAHSIHQRS 311
Db gertgifsnyvrpkdgenvgnas---ksgasnkpkelaqvtsayaas 1053

RESULT 11
AAY57445
ID AAY57445 standard; Protein; 1197 AA.
XX
AC AAY57445;
XX
DF 28-FEB-2000 (first entry)
XX
XX Mouse Ese2 protein sequence.
XX
KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW LH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
OS Mus sp.
XX
PN WO9955728-A2.
XX
```

```
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
XX
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
PI
XX WPI; 2000-052802/04.
DR
DR N-PSDB; AAZ39010, AAZ39011.
XX
XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
XX Claim 33; Page 48; 99pp; English.
XX
XX The present sequence represents mouse Ese2. The present invention
CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
CC variants (Ese - EH-domain and SH3-domain regulator of endocytosis). (I)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) or its (ant)agonists, mimetics, fragments and inactive
CC mutants; (I)-specific antibodies (Ab); sequences antisense to the (I)
CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Ese1 is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor;
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Esp15 complex, then binding dynamitin to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
XX
SQ Sequence 1197 AA;

Query Match 8.78; Score 177.5; DB 21; Length 1197;
Best Local Similarity 23.28; Pred. No. 6.6e-08;
Matches 63; Conservative 50; Mismatches 108; Indels 51; Gaps 10;

Qy 80 WF-----DGQRAAENRQGTITEXCSTLMSLPTKISRCP-HLLDFKVRPDDKLPT 129
Db wfcnyvekvlssekalspkallpptvs--lsatstssqppasvtdyhvnsfslntvt 824

Qy 130 DNQYKKPET-----YLMPPKDGKSTATDITGPIILOVTRAIANYEKTSCSEMA1STGDV 183
Db twgqksaftvtvsgvsgplhggggavenl-----kagaicswtakkenhlnfskhdvi 878

Qy 184 EVVEKSEGWFCQKAKRGWIPASFLPLELSDPD-ETEDPEPNYA----- 227
Db tvleqqen--wfigevhgrgwfpskypvkiipgnevrgrepealyaavtkktpstapvts 937

Qy 228 -----GEPYVAIKAYTAVEGDEVSLLGEAVEVHKLDDGWKDDV---TGYPFPMYLOKS 279
Db taypvgeeyalyssvpegdltftegeellvtqkdgewtgtsigertgifsnyvrpk 997

Qy 280 QGDVYSAQRQIKRGAPRRSSIRNAHSIHQRS 311
Db gertgifsnyvrpkdgenvgnas---ksgasnkpkelaqvtsayaas 1027

RESULT 12
AAY57450
ID AAY57450 standard; Protein; 1658 AA.
```


CC their specificities (e.g. pharmacological activities) can be assessed
XX using the method of the invention.
SQ Sequence 248 AA;

Query Match 8.74; Score 176.5; DB 17; Length 248;
Best Local Similarity 28.04; Pred. No. 7.4e-09;
Matches 59; Conservative 37; Mismatches 68; Indels 47; Gaps 11;

QY 114 LLDFEKVRPDLKLPDQ---TKK-----PENYLMPKD-----GKS 147
DQ 45 lpyssvgpdlftgeeeilvtkdqgwtsgidrgsfygvkpkdgesfgsaks 104
QY 148 TATDITGPIILQTYRAIANYEKTSSEMALSTGDVVEVEKSEGWFCOMKAK-----R 202
DQ 105 gasn-kpetaqvtsa---yyasgseqlslapqqlilkkntsgwggelqargkkrqk 160
QY 203 GWIPASFLPLELSDPEDEPNVAGEP---YVAIKAYTAVEGDEVSLLEGEAVEVEIHK- 258
DQ 161 gwfashvklilgppsseratp---afhpvcqviamydaannedelsfskgqlinvmnk 216
QY 259 LLDGKDD---VTGYFSPMYLQ-KSGQDVSO 285
DQ 217 dpdwwgpeingvtglfpsnyvymtdsdpsq 247

RESULT 14
AA57444
ID AAY57444 standard; Protein: 1214 AA.
AC AAY57444;
XX 28-FEB-2000 (first entry)
DE Mouse Esel protein sequence.
XX Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX Mus sp.
XX WO9955728-A2.
XX 04-NOV-1999.
XX 27-APR-1999; 99WO-CA00375.
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX (HSC-) HSC RES & DEV LP.
XX Egan SE, Wang W, Sengar A;
XX WPI: 2000-052802/04.
XX N-PSDB; AA239008, AA239009.
XX New nucleic acid encoding Esel and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection
XX Claim 14; Page 43; 99pp; English.
XX The present sequence represents mouse Esel. The present invention
CC specifically describes mammalian Esel and 2 proteins (I) and their splice
CC variants (Ese - EH-domain and SH3-domain regulator of endocytosis). (I)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Esp1 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)

CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Esel is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC Antagonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor;
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Esp1 complex, then binding dynam to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
XX Sequence 1214 AA;

Query Match 8.54; Score 172; DB 21; Length 1214;
Best Local Similarity 20.14; Pred. No. 2.3e-07;
Matches 88; Conservative 67; Mismatches 160; Indels 122; Gaps 17;

QY 36 SEKVYVRRFEIYEFHKTLEMFIEAGAINPENRIIPHLPAKWFDC----- 83
DQ 738 svkvyyr--alypfesrshdeitlqpgdlvmvde--sqtgepgwlgelkgktgwfpan 793
QY 84 --ORAAENROGTLTEYCSLMSLP-----TKISRCP-HLLDFEKVRPD 123
DQ 794 yaekipenevtpakpvtldtsapapklalretpaplvtsesepsttpnnwadtsstwp- 852
QY 124 DLKLPDNTQTKKET-----YLMPKDGK-----STATDIT-GPIILQ--- 159
DQ 853 -----ssnekpetdhwtdwaqpsltvpsagqlrqrtsatpataatgsspsvlggqek 906
QY 160 ----TYRAIANYEKTSSEMALSTGDVVEVEKSEGWFCOMKAKRGWIPASFLPLELDS 215
DQ 907 veglqadalyprakrdhnlfnksdvitvlegqdm-wwfgevqgkgwfpkysvkiisg 965
QY 216 P-----DETEDP-----EPNAGBPYVAIKAYTAVEGDEVSLLEGEAVEV 255
DQ 966 pvrktsidgtptespaslkrvaspaakpaipgeefiamytesyesegqdlffqgddvivr 1025
QY 256 IHKLDDGKWK---DDVTGYFSPMYLQ-KSGQDVSOAQROIKRGAPPRRSSIRNAHSIHQR 311
DQ 1026 tkdgdwvtgtvgdksgvfpsnyvrlkdsesgtagktgslgkpketaqviasyaatgpe 1085
QY 312 RKRLSQDAYRNSVRFLOQRROARPG-----PQSPCSPLEEERQTRSKPQAPVPRP 365
DQ 1086 qltla-----pqqlilrkknpqgwwegelqargkkrqigwfpn 1125
QY 366 SADLILNRCSSESTRKL 382
DQ 1126 yvkilspgtskitptel 1142

RESULT 15
AA57449
ID AAY57449 standard; Protein: 1715 AA.
XX AAY57449;
AC AAY57449;
XX 28-FEB-2000 (first entry)
DE Mouse Esel protein sequence.
XX Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX Mus sp.
XX WO9955728-A2.

